European Commission, Research Directorate-General
Directorate C - European Research Area
Universities and Researchers

Feasibility Study

for Creating a European University Data Collection

[Contract No. RTD/C/C4/2009/0233402]

Final Study Report

Document Information Sheet

Deliverable Title	**Final Study Report**
Abstract	This document overviews the EUMIDA Study and its context, objectives, methodology, results and reasoned conclusions – including an outline of the proposed statistical infrastructure. This final version is grounded on the available evidence at the time of delivery from Country level Data Collections 1 & 2, which have been separately delivered as Annexes. It also takes benefit from a variety of comments and contributions received on the previous draft from the Commission services as well as Eurostat.
Authors	Andrea Bonaccorsi (PISA), Tasso Brandt (FRAUNHOFER), Daniela De Filippo (USI), Benedetto Lepori (USI), Francesco Molinari (PISA), Andreas Niederl (JOANNEUM RESEARCH), Ulrich Schmoch (FRAUNHOFER), Torben Schubert (FRAUNHOFER), Stig Slipersaeter (NIFU STEP)

Authorship

The EUMIDA Consortium consists of:

- University of PISA, Facoltà di Ingegneria, Dipartimento Sistemi Elettrici e Automazione, Italy
- FRAUNHOFER – Gesellschaft zur Foerderung der angewandten Forschung e.V, Germany
- JOANNEUM RESEARCH – Forschungsgesellschaft mbH, Austria
- NIFU STEP – Norwegian Institute for Studies in Innovation, Research and Education, Norway
- USI – Università della Svizzera Italiana, Switzerland

Contents

Executive Summary

A. Introduction

The present study (also known as "the EUMIDA project") has laid the foundations for a regular data collection by national statistical institutes on individual higher education institutions[1] in the EU-27 Member States together with Norway and Switzerland. The related contract with the European Commission was signed on 6th July 2009 for a duration of 15 months, thus ending on 5th October 2010.

The project was carried out by an international Consortium composed of:

- University of PISA, Facoltà di Ingegneria, Dipartimento Sistemi Elettrici e Automazione, Italy (Coordinator)
- FRAUNHOFER – Gesellschaft zur Förderung der angewandten Forschung e.V, Germany
- JOANNEUM RESEARCH – Forschungsgesellschaft mbH, Austria
- NIFU STEP – Norwegian Institute for Studies in Innovation, Research and Education, Norway
- USI – Università della Svizzera Italiana, Switzerland

A considerable number of individual experts were involved as National Contact Points for all data collection activities. These are listed in Annex 7.

A dedicated Eurostat Task Force (FESUR) was also set up to provide input and support to the project. This Task Force was composed of statistical representatives from the National Statistical Authorities in around 20 countries together with Eurostat, DG Research and DG Education and Culture. The Consortium would like to thank them all for their invaluable collaboration.

We also wish to acknowledge the timely and effective contributions from our Quality Control Group, consisting of the two Experts Léopold Simar and Giorgio Sirilli.

This document provides an overview of the EUMIDA project, its context, objectives, methodology, results and reasoned conclusions – including an outline of the proposed statistical infrastructure. This final version is based on the available evidence at the time of

[1] Although the title of the project refers to a 'University' data collection, the aim is to cover all "higher education institutions" irrespective of their name and status in the Member States.

delivery from Country level Data Collections 1 & 2, which are delivered separately as Annexes.

The preparation of this Report was led by the University of Pisa with an active contribution from all the EUMIDA Partners.

B. Feasibility of a regular data collection

The main goal of the EUMIDA project was to test the feasibility of a regular data collection of microdata on higher education institutions (HEIs) in all EU-27 Member States plus Norway and Switzerland. The project has reviewed the issues of data availability, confidentiality, and the resources needed for a full-scale exercise. Its main achievement is to have demonstrated that in all countries there actually exists a core set of data that shares the following features:

- it follows the definitions laid down in the UNESCO-OECD-EUROSTAT (UOE) Manual
- it is routinely collected by the National Statistical Authorities (NSAs)
- it does not raise significant confidentiality issues
- it can be disaggregated at the level of individual units in a smooth way.

In more detail, the main results are as follows.

First, in order to explore the feasibility, a preliminary step was to define the perimeter of institutions to be covered. The ToR of the study clarified that the perimeter should involve all institutions delivering degrees at ISCED 6 and ISCED 5a, but also a reasonable set of those delivering ISCED 5b degrees (vocational training). The EUMIDA study adopted an institutional perspective, including in the perimeter those entities that not only deliver degrees on a continuative basis, but also have a substantial autonomy in managing staff and financial resources. This definition excluded a number of small entities, mostly schools associated to industry or professional associations, which deliver ISCED 5b degrees but cannot be considered institutions in the sense outlined above. They may be large in number, but typically enrol a small number of students each.

The study demonstrated that the definition of the perimeter could be completed with large agreement from all NSAs. The study collected data on 2,457 institutions in all countries with the exception of Denmark (which provided only data in Data Collection 2) and France. The total number of HEIs including Denmark and France is estimated at around 2,900. Cases of exclusion have been documented and clarified. Overall, the perimeter includes institutions that enrol 90% of all students enrolled in Europe, as registered by Eurostat. The institutions excluded from the perimeter are typically small schools that deliver ISCED 5b degrees and whose quantitative importance in the higher education landscape is limited. This is a major achievement of the study.

Second, the EUMIDA project investigated whether there are significant obstacles to the collection and publication of data related to individual institutions, in view of a future, regular data collection. It was felt, in fact, that there might be legal obstacles to the publication of data referring to individual units. It turned out that such obstacles are generally-speaking not significant. They are restricted to subsets of institutions in a few countries (typically, private universities) and, in some cases, to financial data. For the overwhelming majority of countries, and basically for all variables in Data Collection 1, there are no obstacles at all. However, in a few cases it's not just legal obstacles but there seems to be a lack of clarity at national level as to whether the data can or should be published.
For Data Collection 2, a number of countries simply do not have comparable data for some of the variables while in other cases, national authorities have not previously published such data at institutional level and therefore need to review their national procedures. However, such outstanding issues do not affect the overall goal of a regular data collection of individual data, to be published in the future. This is a second achievement of the project.

Third, the EUMIDA project carried out two large data collections: one based on a set of core indicators (Data Collection 1) on the entire perimeter (n=2,457), the other based on an extended set of indicators but on a subset of institutions (n=1,364) defined as "research active" (Data Collection 2).
The definition of research active institutions required another stream of conceptual work. The EUMIDA project discarded the approach, which is used elsewhere (e.g. in the Carnegie classification of US higher education institutions), based on the definition of threshold values, such as the absolute number or the intensity of PhD students. The introduction of fixed thresholds is useful for classification purposes, but is inevitably arbitrary from a statistical point of view. Rather, the project adopted a multi-criteria approach, according to which an institution is considered research active if it satisfies at least three criteria out of a list of six. The list of criteria was designed with the explicit goal that any combination of three or more of them would describe an institution that might be sensibly considered as systematically active in research.

Criteria for inclusion have been the following:

- The existence of an official research mandate.
- The existence of research units institutionally recognised (for example on the institutional website).
- The inclusion in the R&D statistics (availability of R&D expenditure data), as sign of institutionalised research activity.
- Awarding doctorates or other ISCED 6 degrees.
- Consideration of research in institutions strategic objectives and plans.

- Regular funding for research projects either from public agencies or from private companies.

Thus, the set of "research active" institutions is much larger of the set of "research intensive" institutions, whatever the definition adopted (e.g. the Carnegie definition). It was considered that in order to describe the landscape of European HEIs the definition of research active was more relevant.
The EUMIDA study has demonstrated that the criteria could be applied in a relatively straightforward way. A few controversial cases are carefully discussed in this Report. A subset of n=1,364 institutions was then identified and constituted the basis for Data Collection 2.
However, the collection of data on research active institutions (Data Collection 2) proved to be much more problematic. This is due to the lack of standardized definitions of some statistical variables (in particular, the breakdown of funding and expenditure by categories) and to the lack of data for many output variables in many countries (e.g. publications, patents, or spinoff companies).

Fourth, the study investigated the cost and effort needed to carry out a regular data collection to be carried out by Eurostat in the near future. After extensive cooperation between NSAs and the network of country experts selected by EUMIDA, it was possible to build up a reasonable estimate of the workload under the assumption of a regular activity of NSAs in data collection and transfer to Eurostat on an individual basis. It turned out that the overall workload is in the order of a few days or weeks per country, with a few exceptions. This implies that the overall goal of a regular data collection should not require significant additional resources for most NSAs.

Overall, these findings suggest that a regular data collection is feasible because data is available, the legal obstacles are not overwhelming, the perimeters of institutions are largely agreed, and the overall effort is within the scope of the current activities of most statistical authorities.

C. Main findings

The EUMIDA study offered also a preliminary analysis of data collected from NSAs. The main points, among many findings, are as follows.

C.1 Perimeter

There are 2,457 institutions identified in Data Collection 1: these constitute the perimeter of higher education institutions in Europe (with the exclusion of Denmark and France for which data was not timely and availlable).
There are 1,364 research active institutions: of these only 850 are also doctorate awarding. This means that a significant portion of research active institutions is found outside the traditional perimeter of universities, i.e. in the domain of non-university research (particularly in countries with dual higher education systems). Bibliometric indicators suggest that the research carried out in the non-university sector is less visible than the one in the university sector.
There are also 1,052 non-research active institutions, most of which are non-doctorate awarding[2]. For a small residual group data is missing and the classification cannot be pursued at the moment.

C.2 Highest degree delivered

In terms of the highest degree delivered, 840 institutions (34.2%) deliver up to the bachelor, 675 (27.5%) up to the master, and 892 (36.3%) up to the doctoral degree, while 2% of data is missing; this means that the higher education landscape is formed by three groups of approximately similar size.

If various descriptors are used to build up clusters and their number is optimized, it turns out that only two clusters emerge (in a slightly different specification, a small third cluster is visible, mostly formed by private institutions). These clusters correspond quite precisely to the ***University model*** (i.e. doctorate awarding, research active institutions: 52.2% of the total) and the ***College model*** (i.e. non doctorate awarding, partly active, partly non active in research: 47.8% of the total).
In the clustering exercise national differences do not matter a great deal. This means that the European landscape, notwithstanding several national specificities, is structurally similar to the landscape of other large countries in which there is a differentiation of educational missions across institutions.

However, while the number of non-doctorate institutions is quite large, almost 80% of students are enrolled in institutions with the right to award the doctorate, while 8.8% are enrolled in the 846 institutions that deliver only bachelor degrees and 12.6% are enrolled in

[2] In addition, if:
- France was included, following the estimate of the perimeter done by experts (see also Table 21)
- The same number of institutions for Denmark identified in DC 2 were included in DC 1 (which is a conservative estimate)

then the overall European perimeter would consist of 2,906 institutions, of which 1,498 being research active.

institutions that deliver up to master degrees. This means that the College model is not yet mature in most European countries, in the sense that it does not capture a significant share of students' preferences. It seems that students prefer to study at universities even if they do not reach the highest degrees, rather than attending non-doctorate institutions.
This finding is also important, insofar as it suggests that the degree of internal differentiation of the higher education system is likely to increase in future. More generally, it points to the issue of the relation between higher education and vocational training in Europe. This issue is attracting lot of attention in the light of the new educational needs of the knowledge society, which cannot be addressed only by university-type institutions.

C.3 Research activity

In total, 39% of research active institutions do not award doctorate degrees. There are large variations between countries.
Particularly there are countries in which around two thirds of research active institutions are represented by institutions that do not grant the doctorate. These are Austria, Belgium, Netherlands and Switzerland. In Germany and Hungary this share exceeds 60%. In all other countries the share is significantly smaller.
This finding is interesting, as it sheds light on the nature and size of the non-university research sector in Europe.

C.4 Doctoral education

At the ISCED 6 level, the core data set covers 531,370 students and 92,631 doctorate degrees awarded. The number of institutions offering a doctorate as the highest degree is 885 - equivalent to 36% of all HEIs. A further 5 HEIs report offering an intermediary ISCED 6 qualification and thus have students at the ISCED 6 level. In total, 890 HEIs have students at the ISCED 6 level. Thereof 850 report being research-active, equivalent to 96% of all HEIs with ISCED 6 students. The remainder comprises art colleges, theological academies, defence universities, or specialised HEIs in management or finance.

In the dataset 870 HEIs provide data for students at the ISCED 6 level. Therein the maximum share of ISCED 6 students within all students is 100%, the minimum 0.1%, the median 3.3%. The distribution of this share is extremely skewed. The HEIs with very high shares of ISCED 6 students are generally quite small (in terms of student numbers) and specialised in fields such as theology, arts, or specific technologies. The HEIs with very low shares are primarily teaching/education-oriented. The intersection with HEIs without ISCED 6 students appears to be diffuse. The "standard" universities can be found in the range between 2% and 8%; this share can then be interpreted in terms of research orientation.

C.5 Internationalization

Looking at the distribution of the share of international ISCED 6 students within all ISCED 6 students, available for 843 HEIs, the range is between 0% and 100%. Very high shares can be observed for small HEIs and for larger HEIs in Switzerland and the United Kingdom. Since there are substantial differences between countries in the definitions used, appropriate comparisons should be made between HEIs of the same country.

C.6 Scientific publications

Publications data is not collected by NSAs at country level. There is not even a definition of what scientific publications are from a statistical point of view. This is an area where further studies are needed.

EUMIDA carried out a feasibility-in-the-feasibility study in order to explore the impact of the use of different bibliometric sources on individual profiles of institutions, namely World of Science and Scopus.

The study showed that there are substantial differences by field and country. A general observation is that the coverage of Humanities in WoS is much broader than in Scopus, but contrariwise the coverage of Agriculture is broader in Scopus. Among the disciplines with a high level of publications, the coverage of the Natural sciences in WoS and Scopus is nearly equal, with a slight advantage for WoS. In Engineering and Medicine, the coverage in Scopus is generally broader.

The analyses by country cover all types of institutions, not only HEIs. Furthermore, they refer to articles. The integration of different document types in the analyses can make a significant difference. It turned out that for 21 of the 29 countries the coverage in Scopus is broader than in WoS, in further 3 cases, the level coverage in Scopus is at least at 95 percent of that in WoS. Only in countries with a low absolute volume of publications is the coverage in WoS better than in Scopus, which can be due to specific preferences of some universities for journals not covered by Scopus.

This general comparison at the country level suggests that bibliometric searches should be performed in Scopus, as the only major disadvantage compared with WoS is the low coverage of the Humanities, but in WoS the figures are also so low that they cannot be used for reliable statistical analyses.

In addition we found, by examining 57 institutions sampled from the larger perimeter, that a future study aimed at cleaning affiliations in bibliometric sources might be feasible with a limited investment.

C.7 Academic patents

Out of the HEIs of the enlarged data set, only 399 report on their patent activities, whereof 195 display at least one application. These latter are primarily from universities in Finland, Italy, Norway, Spain, and the United Kingdom. All these statements refer to the group of applications with the HEI as one of the applicants. Only in 4 cases is data provided for applications with at least one inventor from a HEI. This finding shows that, even if the registration of HEI patent applications is organised in a more systematic way, the information on patent applications with the HEI as origin, but without the HEI as applicant, is generally not available. To conclude, it proved impossible to collect systematic and complete data on patents with a HEI as origin by means of a questionnaire based on data at statistical offices or other public agencies.

C.8 Spinoff companies

In the EUMIDA data set, only 282 HEIs report on spin-off companies; thereof only 105 with at least one company. The reporting is even weaker than in the case of patent applications and cannot be considered a valid source for analysis.

C.9 Funding and expenditure data

Data on funding and expenditure tends to be a weak part of statistical systems. Not only is research funding data at individual level reported for a small number of institutions alone (n=504), but it also suffers from lack of standardization. This is an area where further work is needed.

These are only a sample of the preliminary results that can be derived from the analysis of statistical data available. Section 6 of this Report provides more in-depth results. Extending the analysis to Data Collection 2, and above all combining this data with demographic, social and economic statistics, will provide a large platform for future evaluation.

Section 8 of the Report contains a number of recommendations, which have both a practical, short-term impact, but also some far-reaching implications. These include recommendations concerning:

- The publication of Data Collection 1 used in the EUMIDA project
- The setting up of a regular data collection to be managed by Eurostat
- Statistical capacity building and further methodological work on key variables
- Future feasibility studies.

1. Background and goals of the EUMIDA project

1.1 The debate on European higher education, between the Bologna process and the European Research Area

Higher education is traditionally an important object for analysis and policy making, since the reproduction of professional and political élites and the realisation of the promise of upward social mobility in democratic societies depend to a large extent on its effective working. This is even more so after the emergence of the notions of knowledge society and knowledge economy, which imply that the social and economic value of knowledge production (research) and diffusion (education) are expected to grow.

These issues are debated with particular strength in European countries. Europe has invented the idea of university and has created the institutional framework for the development of the modern university, linking research and education, a model that has been imitated by other countries, *in primis* the United States.

In the last decade, the European Union has warmly embraced the goal of becoming a competitive knowledge-based economy. The main areas of policy considered for supporting this goal, in addition to efforts at the level of Member States, have been the Bologna Process on education and the progressive creation of the European Research Area on research.

The Bologna Process standardised Bachelor and Master degrees, which are awarded on the basis of achieved credit points. This shall result in increased mobility, because in principle the credit points (so-called ECTS-points) can be transferred from one European university to any other (Hunerli and Yildirim, 2009).

After a decade, there is still some heterogeneity in implementation. The large study by Westerheijden et al. (2010) has examined the state of the art of the implementation, suggesting a reasonably fast process. However, there are also cases of persistence in the lack of harmonization. Taking for example the case of Italy and Spain, Garcia-Garcia et al. (2009) argue that Italy implemented the reforms quite quickly via governmental regulations, while Spain had until 2009 not yet enacted it. Furthermore, the effective student mobility via the mechanism of ECTS requires at the very least roughly comparable curricula in the disciplines, which indeed is a much longer and essentially decentralised process. The degree of integration therefore varies considerably from university to university and country to country. Also on the level of harmonisation of curricula there appear to be several deficiencies, as Lobato et al. (2010) argue for the case of medical education[3].

[3] On the other hand, Snelgrove et al. (2009) argue for the case of Italy that reforms in medical sciences have had strong and efficient impact, giving rise to the establishment of student-centred curricula model, which is, essentially, in the minds of the European policy makers.

With regard to the field of research the ERA basically is a fully-fledged transnational research system (Stampfer 2010), including sources for research funds (e.g. the Framework Programmes), the organisational infrastructure and intermediary actors (e.g. the European Research Council) as well as research institutions (e.g. Joint Research Centres). The ERA exists in parallel to the national science system and it is in a certain way a research system that overarches the national systems of the member states. Its main idea is to foster mobility and cooperation between national science systems. In consequence, also any national research organisation can be deemed to be part of the ERA.

At the same time, higher education is often described as one of the sources of difficulty in achieving the goals of knowledge based economy and society. Governments, political élites and analysts have repeatedly argued that the European tradition of higher education does not match the new ambitions.

This summary statement has been fuelled by the publication, since 2003, of international university league tables or rankings. Relatively speaking, European universities are not ranked high, particularly in tables based mainly on research output. From a methodological point of view, rankings are highly disputable (van Raan, 2007; Dill and Soo, 2005; Frey and Rost, 2008; Harvey, 2008), and can be extremely misleading. The Berlin principles have been proposed to warn against counterproductive use of rankings (CHE et al., 2006). At the same time, it is difficult to obscure the fact that their publications have catalyzed the policy discussion.

Several commentators have suggested an agenda of radical reform, largely based on the relatively poor position of European universities in international rankings (Jacobs and van der Ploeg, 2006; Thissen and Ederveen, 2006; van der Ploeg and Veugelers, 2008; Aghion et al., 2008).

And also the European Commission, while not formally recognising the value of rankings, has promoted a sustainable pathway of innovation and modernisation in a number of key official policy documents[4], aimed at fully harnessing the potential of EU higher education system in support to the deployment of the renovated Lisbon agenda. For example, the 2007 Council conclusions call on Member States to "promote excellence in higher education and research by developing institutions and networks able to compete internationally and to contribute to attracting to Europe the best talent", as well as to provide "these institutions with the autonomy to develop their full potential".

This idea of streamlining by comparing performance is not accepted without controversy, however. Universities themselves have argued against a simplistic use of rankings, through their representative Association (EUA, 2005a; 2005b; 2007), while also the League of Research Universities produced a position paper in which the overall approach was criticized (LERU, 2006 & 2010). Specialised literature on higher education has also warned against the

[4] In particular, one can mention the Modernisation Agenda for Universities (2006 – full text available from http://ec.europa.eu/euraxess/pdf/COM(2006)_208.pdf) and the subsequent Nov 2007 Council conclusions (see http://www.consilium.europa.eu/ueDocs/cms_Data/docs/pressData/en/intm/97237.pdf).

reduction of the diversity of European institutions along a single dimension, i.e. research excellence (Kyvik 2004 & 2009; van Vught 2004; Huisman *et al*. 2007). In particular, Huisman, Meek and Wood (2007) and Kyvik (2009) have insisted that policy making should clearly appreciate the dimensions of diversity of higher education institutions.

Consequently, the notion of multi-dimensional ranking has been proposed, giving origin to an experimental mapping exercise (van Vught *et al*. 2008) and an on-going large-scale exercise (U-Map, see van Vught et al. 2010). According to the U-Map exercise, there are **six dimensions of diversity**:

(a) Teaching and learning profile
(b) Student profile
(c) Research involvement
(d) Involvement in knowledge exchange
(e) Regional engagement
(f) International orientation.

Following these dimensions, a set of 23 indicators has been developed, on which a large-scale survey has been launched among universities that volunteered to participate to the pilot phase. This exercise, as well as the subsequent U-Multirank[5] launched by the European Commission, has been systematically taken into account in the design and implementation of EUMIDA.

Before entering into the details of the debate, it is then of utmost importance to develop a full-scale analysis of **diversity**, or heterogeneity, of the European higher education landscape. This will be done in the remainder of this Introduction.

1.2 Diversity in European higher education: (a) educational dimension

According to the OECD, diversity 'implies that distinct courses or institutions serve distinct objectives, receiving and responding to distinct streams of students' (OECD 2008, pp. 40-41). In the field of higher education, this definition of diversity has, in turn, several dimensions.

One is the highest degree delivered, or the number of years of formal education offered. Given the standardisation introduced by the Bologna process, this dimension is easily captured by a tripartite progressive structure at the level of courses: Bachelor, Master and Doctorate. This is not easily translated into a classification of institutions, however, because the attribution of activities is asymmetric: institutions accredited for delivering Doctorates

[5] http://www.u-multirank.eu/

are also permitted (and usually do) to deliver Bachelor and Master degrees, while the opposite is not true. In addition, there is some continuity between Bachelor and Master in the organizational practices, so that in reality the most important distinction is between those institutions with the Doctorate or without.

The second dimension refers to the role of higher education institutions with respect to vocational training. This is defined as the professional training aimed at achieving the accreditation to carry out specific job categories, which takes place with courses of three years or less. Vocational training is allocated to separated institutions in some countries, while is the mandate of universities in others, alongside longer curricula. According to Kyvik (2004) and Scott (1995), European higher education systems fall into five groups: university-dominated, dual, binary, unified, and stratified. In university-dominated systems (basically, just Italy) there is no differentiation since all post-secondary training is confined to universities. Dual and binary systems allocate university education and vocational training to separate institutions, while unified systems (UK, Spain) have absorbed vocational training institutions within universities, amalgamating previously separate experiences. Finally, a stratified system does not exist in Europe in the pure form of the US system, although some of its characteristics are present in the French system. For classification purposes, it would be possible to use the same variable as above (with or without doctorate), but the qualitative interpretation is largely different according to the national context.

Finally, a third dimension of diversity alongside the educational mission refers to the subject mix. Teichler (1988 & 2005) defines horizontal diversity in terms of the mix of subjects taught. We know from other studies (Filippini and Lepori 2007; Lepori, Probst and Baschung 2010) that many structural characteristics differ across disciplines, so that comparing institutions with different subject mixes without controlling for the differences may be dangerous. While the previous dimensions are somewhat fixed at national level, giving origin to categories of legally separated institutions (at least in the short term), the choice of subject mix is, to some extent, the result of a match between the strategic decision of individual institutions and the government policies for accreditation.

These dimensions of diversity refer to the (a) teaching and learning profile and (b) student profile dimensions of the U-Map.

Luckily enough, the EUMIDA project demonstrated the feasibility of collecting this data on a regular basis, providing a robust empirical base for examining diversity in the higher education landscape. This is not the case, however, for other dimensions of diversity, which are introduced in the following.

1.3 Diversity in European higher education: (b) research dimension

Higher education institutions differ quite systematically along the research dimension. However, in this case there is considerable difficulty in mapping diversity, because the institutional system of most European countries does not offer any legally or administratively enforced categorization of units according to research. To make an example, it is clear that in the UK system there is a large difference in the research profile among universities, but all of them belong to the same category. Thus, we would expect that most universities that were originally established as polytechnics were less engaged in research than old universities, but there is no official labelling to be used for categorization. This is a sharp difference with respect to diversity in educational dimension, since the latter is formally enforced into different categories (highest degree, field of education). In the educational dimension there are discrete categories to be used for mapping diversity, while in research there is rather a multi-dimensional continuous variable (or vector), without natural threshold or cut-off points.

The U-Map exercise suggested a dimension labelled Research involvement (peer reviewed publications, expenditure in research, doctorate production). This dimension is, at the same time, the one most examined and the most controversial. It is largely examined because one of the indicators of research involvement is easily traceable, namely scientific publications. Indeed, starting with the pioneering works in bibliometrics and scientometrics and the storage of publication data in electronic form in the 1960s, scientific publications have formed an object of investigation per se. More recently, the attribution of scientific publications to universities (affiliations) has been improved, after careful work of disambiguation of affiliation labels, as well as the improvement of the treatment of affiliation names in scientific journals. Consequently, it has become easier to allocate objective data from electronic sources to institutional affiliations, building up publication statistics at the level of universities. The next step has been to compile world rankings of universities, an exercise that became popular since 2003 and is now accepted practice. All these data production activities have been taking place outside the boundaries of official national statistics.

Based on these indicators, what we now have is a complete ranking of top 200 European universities, based on ISI Web of Science publications, as well as the list of European universities included in the ranking of top 200 or 500 world universities. It is not our goal here to enter into a discussion of these indicators. We simply raise a few questions: What proportion of the European higher education landscape is captured in these rankings? What lies below the line? And how could we address the issue of describing and measuring the research activities of those universities (or other higher education institutions) that do produce research, but do not reach the level of visibility needed to be included in the ranking, or produce research which is not easily captured by Web of Science indicators?

As it is quite clear, measurement of the research dimension is the most controversial issue (Schmoch and Schubert 2009a; Schmoch et al. 2010). To start with, there is no accepted statistical definition of what a scientific publication is. Furthermore, there is widespread recognition of the fact that different scientific disciplines have largely different practices for publication. As the Expert Group on *Assessment of University Based Research (AUBR)* recommended, there should not be a unique methodology for measuring publication output.
A related issue depends on the distortion introduced in the analysis when a single source of bibliometric indicators is used. For example, it is recognized in the literature that Web of Science is an appropriate source for Natural Sciences and the Biomedical field, while it is not adequately representative of Engineering, Social Sciences, and Humanities.
Even more intriguing is the consideration that there exist outputs of research which are not publications, but, for example, software, prototypes, drawings or exhibitions (see the next Table). Furthermore, one might consider legitimate outputs of research also the organization of conferences, the editorial work for refereed journals, the election in boards of scientific societies, and the like. While many universities and also some governments take this sub-dimension into account, it is not clear at all how a statistical system could define and track these elements.

Table 1. Overview of research outputs

	Natural sciences	Life sciences	Engineering sciences	Social sciences and humanities	Arts
Journal Article	X	X		X	X
Conference Proceedings			X		
Book chapters				X	
Monographs/Books				X	
Artefacts					X
Prototypes			X		

Source: DG Research. Expert Group on *Assessment of University-based research* (AUBR). August 2009.

These difficulties notwithstanding, it is clear that large diversity can be found across institutions in terms of intensity of activity, volume and quality of output, with respect to research. But in order to map these dimensions we cannot rely, for the time being, on

statistical indicators. We will devote Chapter 3 to a thorough examination of this problem, offering a feasibility study on the potential for extending the cleaning of affiliations from the Scopus dataset to all EUMIDA institutions.

The problematic status of these indicators from the perspective of a statistical system is even more evident if we try to allocate indicators to institutions.

First of all, not all institutions devoted to higher education are also active in research. Some of them do not have a research mandate in their institutional definition so that, for example, are not eligible for receiving research grants. Others do have an institutional legitimisation, but in practice do not carry out formal research activity, due to lack of resources or organization. The EUMIDA project developed a definition of research activity that proved quite robust to qualitative investigation (see Table 4 and Chapter 3 of this Report). Thus a very simple, yet extremely informative, indicator has the form of a dummy variable: research active/ not research active.

Second, when coming to research expenditure, there is considerable disagreement on the way in which the public expenditure in higher education can be allocated to research at the level of individual university. As we will discuss in detail in Chapter 2, there is neither a universally accepted standard for research expenditure in public accounting or in financial reporting to governments, nor a reliable statistical practice for turning official data on expenditure into a share on research expenditure.

Given these difficulties, it is not surprising that one of the most robust indicators of research is to be found at the boundaries with educational activity. In fact, the number of Doctorate degrees awarded gives a non-exclusive indicator of research activity. There is not a one-to-one correspondence between Doctorate granting institutions and research active ones, since the latter include also the so-called non-university research sector. This is formed by those institutions delivering only Master and Bachelor degrees, which, however, engage themselves into research activities, typically of applied type. Having said that, it is also clear that granting Doctorate degrees is an institutional activity that includes both educational and research dimensions. An appropriate measure that takes into account differences in size is the ratio between the Number of ISCED 6 students and the Total number of students. In addition, the personnel resources allocated to PhD students give an indirect indicator of research activity. We do not know exactly which part of their budget time is allocated to graduate education (if any). However, a simple ratio between the Number of academic staff and the Number of ISCED 6 students may convey an approximation of the importance of graduate education.

1.4 Diversity in European higher education: (c) knowledge exchange

Universities produce new and valid knowledge. Through education, they diffuse knowledge to students, and through their professional and social life, to society and economy as a

whole. Through publication of scientific results, academic researchers diffuse knowledge mainly to other researchers, who use the results as intermediate products for their own research, in a cumulative process. The notion of third mission, or knowledge exchange, has been introduced to mean those activities through which universities transfer knowledge in a less mediated way to users. These include companies, but also actors in society, such as public administration, civic associations, or the non-profit sector.

This activity has been labelled "third mission", in order to emphasize the fact that it is recognized as an institutionally legitimate dimension of involvement.

This dimension has been largely explored in the last two decades, but mainly along the boundaries between universities and firms, i.e. university-academia collaboration (Gulbranson and Slipersaeter 2007; Holi, Vickramasinghe and van Leuwen 2008).

There are claims that the interactions between scientific and non-scientific institutions have grown largely in the last part of 20th century, particularly along the dimension of academia-industry relations. Indicators of such a growth have been identified in the increase of references to academic papers in patents (Narin and Olivastro 1992; Grupp 1992; Narin, Hamilton and Olivastro 1997; Jaffe, Trajtenberg and Henderson 1993; Gittelman and Kogut, 2003), of citations to academic papers in publications of industry researchers (Adams and Clemmons, 2006), of academic patents (Henderson, Jaffe and Trajtenberg, 1998) and licensing (Thursby and Kemp 1998; Jensen and Thursby 2001; Thursby, Jensen and Thursby 2001; Thursby and Thursby 2002; Zucker, Darby and Armstrong 2002; Owen-Smith and Powell 2003), of funding and research collaborations between industry and university (Guston and Keniston 1994; Agrawal 2001; Friedman and Silberman 2003; Laursen and Salter 2004; Link and Siegel 2005; Bercovitz and Feldman 2007), of co-authorship of papers between academic and industry researchers (Van Looy *et al.* 2004), of co-invention of patents (Bonaccorsi and Thoma 2007) and of academic entrepreneurs (Di Gregorio and Shane 2003; Rothaermel, Agung and Jiang 2007).

An influential stream of literature has stressed that universities do not simply "transfer" knowledge, but are actively engaged into two-way interactions with industry and government (Etzkowitz, 1998 & 2003; Etzkowitz and Leydesdorff, 1998 & 2000).

The debate on knowledge exchange activities has also taken a distinctive European dimension. According to the Green Paper on Innovation in 1995, and after that according to several official documents of the European Commission in the subsequent decade, European universities are less active than US universities in commercialization of research and technology transfer. This may be due to differences in the degree of professionalization of boundary personnel. Thursby and Kemp (1998) and Thursby, Jensen and Thursby (2001) have shown that the successful commercialization of research in US universities is the result of a strong professionalization of boundary roles and of allocation of dedicated organizational resources. However, other studies suggest a more subtle picture. For example, Conti and Gaulé (2009) compared the activities of Technology Transfer Offices of European and US universities and concluded that the main difference is not in the quantity

of licenses, but in the revenue generated. This is largely due to the lack of professional resources at TTO that come from the managerial career.

As we will see later, in the case of patents Lissoni *et al.* (2008) discovered that a significant portion of patents resulting from inventions of academicians in European universities do not follow the formal route of official university patents, but are assigned to a variety of other actors. If the inventive productivity of European scientists were computed by taking into account both official and non-official academic patents, it would not be lower than the one in USA.

With respect to spinoff companies, Wright *et al.* (2008) have shown that European companies have a tendency to grow less than US ones. The extent of this depends on differences in the academic environment, or the governance and organization of universities, or financial markets, or labour markets, is still controversial.

Although these issues are the subject of a large dedicated literature, there is no large scale statistical coverage of them.

The European Commission has invested large resources in an effort to measure these phenomena, as well as to improve the circulation of knowledge and practices. Based on an extensive analysis of all survey carried out in European countries at the level of university and PROs, the Expert Group on *Knowledge Transfer Metrics* (European Commission, 2009)[6] suggested the adoption of the following indicators (core performance indicators for the PROs served by the Knowledge Transfer Office):

- Research agreements
- Invention disclosures
- Patent applications
- Patent grants
- Licences executed
- License income earned
- Spin-offs established

When reporting the coverage of information collected in the various surveys, it turned out that information is available at individual level for slightly more than 200 universities, 211 out of estimated 918 (23%). When considering the aggregate level, 415 universities participated to the surveys, but the majority did not authorize the disclosure of individual data.

Another project sponsored by the European Commission, E3M, has suggested that the following dimensions of third mission should be considered: (a) Continuing Education; (b) Technology Transfer and Innovation; (c) Social Engagement (Montesinos *et al.* 2008). While this suggestion is useful in order to balance third mission activities related to the economy and to society, it is not easy to devise how to build up comprehensive indicators.

[6] http://ec.europa.eu/invest-in-research/pdf/download_en/knowledge_transfer_web.pdf

In addition, CWTS has published a ranking of the top 200 European universities mostly inclined to collaborate with industry exchange (University-Industry Research Cooperation Scoreboard 2009-2010[7]). This uses co-publications of scientific publications by authors affiliated to universities and authors affiliated to firms as an indicator of knowledge (Tijssen, van Leeuwen and van Wijk 2009).

What these contributions point to is a need for better measurement of knowledge exchange indicators. As already mentioned, we will devote Chapter 3 to a full scale discussion of available indicators, and to a feasibility exercise on new indicators that might be compatible with the existing statistical infrastructure.

1.5 Diversity in European higher education: (d) international and regional orientation

Higher education is increasingly becoming an internationally open activity. Students are encouraged to move between countries and harmonisation policies have been created with the purpose of facilitating short term and long term mobility of students.

On the other hand, in order to attract students from abroad, an institution must have some distinctive features that justify the extra cost. In addition, it must offer courses in foreign languages (usually in English) and/or facilities for learning the domestic language, as well as special facilities to socialize students. In this perspective, the proportion of students that come from other countries may be considered an indicator of attractiveness, or even an indicator of strategic orientation towards international competition (Brandeburg and Federkell 2007; Nuffic 2010).

Similar considerations are valid for PhD students, whose international mobility is larger.

From a related but different perspective, institutions that attract academic staff from abroad are typically internationally oriented, offering state of the art research facilities, as well as interesting teaching opportunities.

From these considerations and on the basis of the available data, the following indicators of international orientation can be developed:

(i) Share of international students ISCED 5
(ii) Share of international students ISCED 6
(iii) Share of academic staff coming from abroad.

These indicators will be able to offer a full picture of the dimension labelled "International orientation" by U-Map.

Unfortunately, there is no statistical data on which it would be possible to build up reliable indicators of "Regional engagement" (see Section 2.4.6 below).

[7] http://www.socialsciences.leiden.edu/cwts/products-services/scoreboard

1.6 Convergence vs path dependency in the dynamics of differentiation[8]

The discussion above has shown that there are several dimensions along which we may expect to find significant heterogeneity across HEIs in Europe. EUMIDA data collection will provide robust empirical evidence on this subject.

At the same time, there is another controversial issue in the debate on European higher education - whether diversity is persistent over time, or we witness convergence towards a common model: "*The question to be discussed is the extent to which the various countries converge to a common structural model for the organization of higher education – either a binary system which is the most common model today, or a unified but hierarchical system as in the United Kingdom*" (Kyvik 2004, 393). In fact, Meek *et al.* (1996) suggested that there are two contrasting perspectives on the evolution of university models: convergence or path dependence.

1.6.1 The convergence thesis

The convergence thesis predicts that one of the following models will prevail: (a) universities and vocational training institutions will be decreed by law to be separate institutions; (b) universities will dominate the higher education system: they will absorb vocational training institutions and converge to a unitary system characterized by internal hierarchisation.

The binary system is the most popular in Europe: according to Kyvik (2004), it has been adopted by the Netherlands, Germany, Belgium, Sweden, Norway, Ireland, Greece, Portugal, Denmark, Finland and Switzerland. While its stability is assured by strong institutional separation, there is evidence that the non-university sector is increasingly adding research activities to its mandate, which is producing some overlaps with universities (academic drift). The pattern of university dominance is clearly visible in the UK, and in the Anglo-Saxon countries more generally. There are several arguments supporting the view that a unitary system eventually will be associated with hierarchisation. First, non-university institutions are driven to adopt university values and norms by powerful imitation and social pressures. This phenomenon, known as academic drift, results in greater homogeneity (Neave 1983)[9]. Second, since vocational training is subsumed within the university system, there is a need

[8] This Section is based on Daraio et al. (2010), with permission.

[9] Counter strategies in a number of European countries, aimed at preserving an elitist element within the higher education system through the creation of a binary or stratified system, have failed. The idea that an effective formal division can be established and maintained, between institutions that focus on pure research and those that take a more utilitarian approach to knowledge production, in order to protect the former against 'external influences', has so far been unsuccessful. Whilst non-university institutions have tried to become research institutions, research universities have never given up more formal, applied research and vocationally-oriented education programmes. Experiments at formal divides have broken down for the reason that attempts to isolate the 'scientific' core have been based on premises (the aim of preserving elite status) that underestimate the forces of 'academic' and 'applied' drift within higher education. In other words, it illustrates the way that the 'scientific core' expands, while at the same becoming integrated with 'social', more utilitarian demands and needs in new settings (Bleiklie 2003).

for a clear hierarchisation among previously university institutions. The absorption of vocational training institutions within the university system (by law or through mergers), is promoting demand for some hierarchy in funding. Third, there is a more general trend towards new forms of integration of teaching and research within the so-called Mode 2 production of knowledge (Gibbons *et al.* 1994), which is demanding that all higher education institutions should be research active.

In turn, a test of convergence can be understood in very different ways. The first and commonly implicitly used view looks at organisational structures and defines convergence in terms of institutional convergence, that is, essentially in terms of harmonisation of governance patterns (Amaral, Meek and Larsen 2003). A common argument is that HEIs converge because they are increasingly subject to similar pressures from the State. Of particular importance is the New Public Management (NPM) Paradigm, which, like a wave, has swept over most Western countries, implying harmonisation in the organisation of the national public science systems (e.g. de Boer *et al.* 2007; Meyer 2007; Frolich 2005; Smith 2004). In contrast to the organisational structures that were common in many continental European countries such as Austria, France and Germany, the spirit of NPM consists of two pillars (Braun and Merrien 1999; Schimank 2007a & 2007b). First, the decision-making competencies of the state authorities are reduced, especially at the purely operative level, leaving much greater steering autonomy to the researchers. Second, the internal hierarchy is strengthened, that is to say, the management authorities (the deans and the university presidents) gain much greater power over the researchers. While the debate on institutional convergence is still open, we will not address it because no indicators are available on a statistical basis.

Another view of convergence looks, on the contrary, at activity profiles rather than organisational structures, which implies that this view would define convergence in terms of the degree to which universities in different countries are functional equivalents.

From a theoretical point of view, it is not clearly discernable, which is the right way to analyse the problem at hand, i.e. whether the organisational or the activity dimension is the more correct one. In any case, for empirical work this question quite likely is of minor importance, because a clear distinction between the organisation and behaviour will not be feasible in any case. This is since institutionally very different HEIs will probably also have quite different activity profiles. Statistically speaking, both dimensions will be highly correlated.

Since data in EUMIDA Project primarily relates to activities profiles of HEIs and not to the governance setting, we will focus, at least in the empirical section, on the view that defines similarities and dissimilarities between HEIs in terms of what they do rather than their organisation.

1.6.2 The path dependency thesis

On the other hand, the institutional, path dependency thesis sees significant resilience in the vocational training sector and that a pattern of hierarchisation pattern is far from established. In this view, a variety of solutions have emerged, based on national and institutional path dependency. According to this perspective, we do not see convergence, because under an apparently similar institutional structure a variety of solutions are emerging (Musselin 1999; Bleiklie 2001; Kogan *et al.* 2000). The trend, beginning to emerge in Austria, Germany, the Netherlands and Switzerland, for the non-university sector to engage in research, is proof that the structural dynamics of academic drift may survive even in binary systems where institutional separation is legally established.

In dynamic terms, the debate on convergence or path dependence is at the core of the related issue of marketisation. In most OECD countries, particularly the USA, the UK, Ireland, Australia, Canada and Israel, there has been a drive towards an increased share of the university budget from private sources. These include: student fees ('user pays' principle); contract research for industry; contracts, grants and donations from private foundations; and sale of licences or products. The impact of more private sources of funding on the autonomy, long-term independence, equity of access and cultural vitality of universities is the subject of passionate debate (Meek 2000). A common theme in this debate is whether marketisation will promote reduced horizontal diversity and increased hierarchisation.

1.7 Debating without tabulating?

This debate is remarkable because it takes place, even in the case of empirical analyses, without knowing which is the universe on which statements can be sensibly be formulated. At the current state of statistical sources, most statements refer either to aggregated data at national level, such as the number of students or government expenditure, or to samples of universities accepting to fill questionnaires. Aggregated data is of course extremely valuable, but cannot be used to examine the issue of internal diversity of institutions.

On the other hand, samples drawn from an unknown universe are by definition non-statistical samples, and then do not allow in any meaningful sense inferential exercises. Even worse, we know from other fields (e.g. economics of science) that some of the variables of interest, such as scientific productivity, are not distributed in a Gaussian way, and in some cases are extremely skewed. This means that knowledge from observed cases not only does not provide information about the mean of the sample, but can be seriously misleading.

In some sense, this situation is similar to a government that must adopt an export policy without knowing which industrial sectors export more, or a Central Bank that must regulate

the financial sector with no clue about the financial products offered by intermediaries, or about the relative efficiency of small or large banks.

What is needed is an official census of institutions, which may establish the statistical foundation for aggregation and disaggregation of data, as well as for inferential exercises based on sampling. Data must be tabulated at the least possible level of aggregation, in order to allow larger degrees of freedom to users for analysis.

2. Methodological issues

2.1 Introduction

This Chapter presents the basic conceptual foundations of the EUMIDA approach to institutional-level indicators for higher education institutions (HEIs), as well as the main definitions concerning the perimeter to be considered and the data to be collected. Further, we outline data collection and management procedures, as well as quality control for the two sets of data to be collected in the framework of EUMIDA:

- a *core set of data* to allow a broader characterization of higher education institutions throughout Europe using a small number of variables. This data should be collected for the extended perimeter of HEIs in Europe.
- a full set of data allowing a more in-depth analysis of inputs and outputs of HEIs, including also a detailed breakdown by scientific fields. This data should be collected only for a restricted perimeter of research-active HEIs.

The focus of the Chapter is on the conceptual rationale for the choices made in designing the statistical system, as well as on the basic definitions and methodological problems that have emerged. For full detail of definitions and implementation guidelines the reader should refer to the EUMIDA Handbook annexed to this Report. For more detailed information concerning data availability and quality problems in individual countries the reader should refer to Chapters 4 and 5 of this Report.

2.2 The EUMIDA conceptual framework

A statistical system on European Higher Education Institutions needs to be built on a clear coherent conceptual framework, which allows the identification of the relevant concepts and variables to be measured and thus the construction of a coherent set of definitions. Moreover, the goals of data collection need to be clearly specified, since this will drive the choice of variables and make it easier to devise some reasonable simplifications, when needed.

From the sociology of S&T indicators (Barré 2004; Godin 2005), it is well know that there is no "objective" choice on each of these issues. Choices concerning indicators are driven by different forces – concepts and theories driving our understanding of higher education, political will and social norms, existing statistical practices and availability of data. A major challenge in constructing the EUMIDA system was to find a reasonable balance between these requirements and to get a system of indicators which is acceptable for the interested stakeholders, feasible in terms of data collection and sufficiently close to the existing practices of national statistical institutes and, last but not least, which can stand in terms of conceptual and methodological rigour.

2.2.1 Basic assumptions

The EUMIDA data collection is based on three main assumptions (Bonaccorsi and Daraio 2007):

- First, an understanding of higher education institutions as formal organizations which are able to act strategically and to choose actively their positioning in the research and higher education areas. We acknowledge all the complexities of considering these institutions as organisations with well-defined boundaries and objectives and strategic capabilities, when faced to strong regulatory control from the State and to the traditional autonomy of academics (Musselin 2007). However a consistent literature in the field has shown that increasingly higher education institutions are granted some degree of autonomy from the State, while incentives are being created to define more focused profiles. Hence, the choice in EUMIDA to adopt individual higher education institutions as the most relevant level of analysis, rather than educational programmes or whole economic sectors at national level.
- Secondly, we consider higher education institutions as multi-input multi-output organisations which use sets of inputs – financial resources, human resources, infrastructure – to produce multiple sets of outputs, including research output, educational outputs, transfer activities towards society and economy. We consider that production processes for these different outputs cannot be easily separated and hence the choice to characterize the whole set of inputs and outputs at the level of the whole institution instead of looking to individual activities as it is current for example in the R&D statistics.
- Thirdly, we have a strong feeling that the European higher education system is characterized by a large diversity of status, size, orientation towards research and education and that this diversity is a relevant asset in order to fulfill the multiple functions of a modern higher education system and to adapt to the high diversity of context across Europe (van Vught *et al.* 2008). Hence, we share a strong focus on characterizing profiles rather than on ranking HEIs along a single dimension.

Thus, the basic conceptual scheme of EUMIDA considers an higher education institution as our unit of analysis, characterized by a set on inputs and outputs, as well as internal processes and embedded in a broader environment, related to state regulation, funding sources, potential pool of students (see Figure 1).

We acknowledge that this environment widely differs from country to country, as shown by comparative studies of higher education systems (Amaral *et al.* 2002; Paradeise *et al.* 2009). While this cannot be covered by EUMIDA data, contextualisation to each national system has to be carefully taken into account in order to interpret the data.

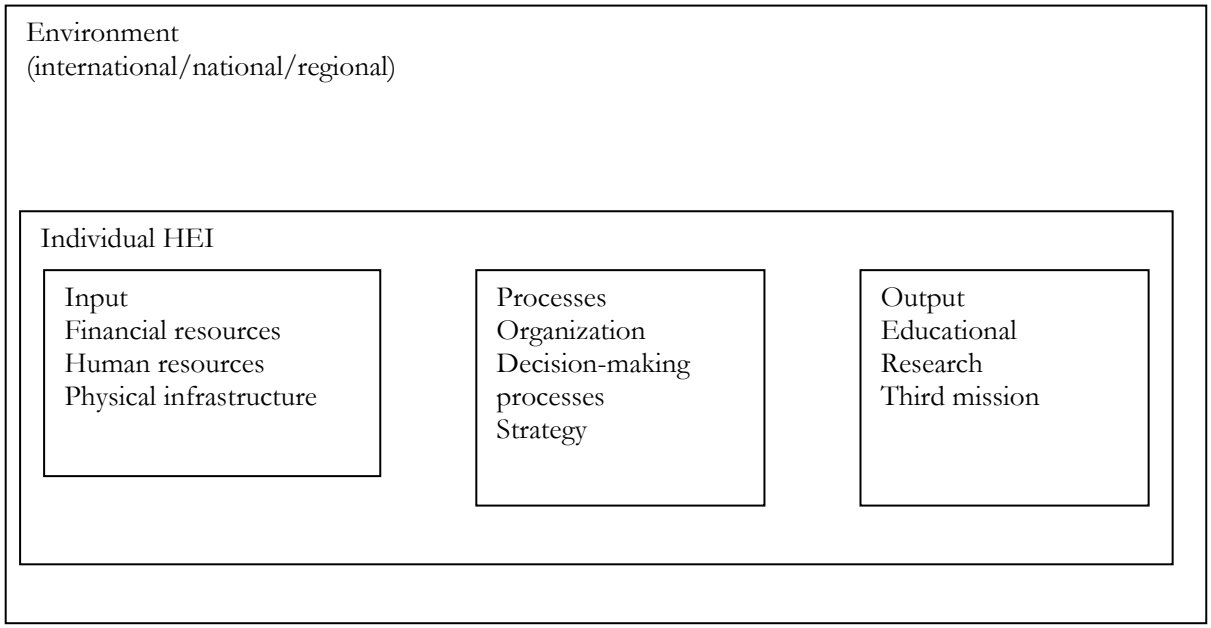


Figure 1. A framework for HEI indicators

2.2.2 Core set of data

The definition of the core set of data is driven by the requirement of characterising the diversity of higher education institutions in Europe. This leads to the identification of a set of dimensions relevant for the HEI activity profile and to a set of indicators, which are able to characterize their relevance.

This approach thus follows closely from the work done in the European projects on characterizing higher education institutions (CEHEI and later U-Map - http://www.u-map.eu) coordinated by the Center for Higher Education Policy Studies at the University of Twente (CHEPS) in the lifelong learning programme of the European commission (van Vught *et al.* 2008). This effort has tried systematically to identify dimensions and indicators for characterization working on a small sample of HEIs across Europe through interviews to relevant stakeholders and clustering of the collected data.

The U-Map project identified six relevant dimensions for characterising higher education institutions in Europe, as well as a number of indicators for each dimension (see Table 1).

Table 2. U-Map dimensions and indicators

Dimension	Indicators
Educational profile	Orientation of degree; Subject areas covered; Degree level focus; Expenditure on teaching
Student profile	Mature or adult learners; Students enrolled (headcount); Part-time students; Students enrolled in distance learning programs
Research involvement	Expenditure on research; Peer reviewed publications; Doctorate production
Knowledge exchange	Cultural activities; Income from knowledge exchange activities; Patent applications filed; Start up firms
International orientation	Foreign degree seeking students; Importance of international sources of income; Students sent out in European and other international exchange programs; Incoming students in European and other international exchange programs; Non-national teaching and research staff
Regional engagement	First year bachelor students from the region; Importance of local/regional income sources; Graduates working in the region

Source: U-Map project, http://www.u-map.eu/

For the purposes of EUMIDA, we take from this work the definition of the relevant dimensions, as well as some suggestions for the indicators to be used. However, while U-Map was designed to be implemented through a survey of the involved institutions, EUMIDA is based on data available in national statistical systems, as shown by preliminary mapping work done in EUMIDA (see Chapter 4 for further details on availability) and thus is a compromise between coverage of the relevant dimensions and feasibility. We discuss in detail the chosen variables and definitions later in this Chapter.

2.2.3 Extended set of data

The extended set of data applies to research active institutions. The definition of research active institutions requires conceptual clarification. The EUMIDA study discarded the approach that is used elsewhere (e.g. in the Carnegie classification of US higher education institutions), based on the definition of threshold values, such as the absolute number or the intensity of PhD students. The introduction of fixed thresholds is useful for classification purposes, but is inevitably arbitrary from a statistical point of view. Rather, the project adopted a multi-criteria approach, according to which an institution is considered research active if it satisfies at least three criteria out of a list of six. The list of criteria was designed with the explicit goal that any combination of three or more of them would describe an institution that might be sensibly considered as systematically active in research.

The extended set of data collected by EUMIDA aims at characterising more completely the set of inputs and outputs of higher education institutions, providing more precise quantitative data which are also disaggregated by scientific field. Thus, this requires a full

characterisation of a number of inputs and outputs. We thus choose the following dimensions:

a) For inputs: human resources (personnel), finances, physical infrastructure, students;
b) For outputs: educational production, research production, third-mission.

These types are further disaggregated by relevant subtypes – for example for staff between academic staff and non-academic staff -, by level of quality, especially for outputs, and by subject domains to allow a more fine-grained analysis of subject specialisation of HEIs. These categories need to be based on theories and studies in higher education, which enable to identify the most relevant characteristics for comparing HEIs, but are also largely built on existing classifications in national statistical systems.

We notice that this approach largely disregards the internal process to HEIs and their internal organization and just observes their effects in terms of the realised mixes of inputs and outputs. This is driven also by following considerations:

- Firstly, in policy terms sufficient information can be provided through characterisation of inputs and output, while it is not the task of public policies to intervene in internal mechanisms of HEIs (following a widely accepted notion of institutional autonomy).
- Secondly, information on internal governance and organization is difficult to collect and, especially, to bring back to standardized categories as required by official statistics. Most of it is likely to be available only in form of qualitative descriptions. Thus, this kind of information could be eventually integrated with the EUMIDA data for individual HEIs or for small samples, but it is questionable if it is meaningful and feasible to collect it systematically in statistical systems. Thus, in the framework of EUMIDA, data collection is limited to the core descriptors included in the core set of variables.

However, a set of basic descriptors of HEIs, including for example legal status, institutional type, the presence of a university hospital is already included in the core set of data and can be exploited also for analysing also the full set of data.
A useful starting point for this discussion is the list of variables adopted for the PRIME-AQUAMETH project, which is presented in the next Table (see Bonaccorsi *et al.* 2007b).

Table 3. Main data in the AQUAMETH 11+ 2 database

Area	Categories
General information	Year of foundation; Region (NUTS); Type (university, technical college etc); Governance (public, private); University hospital (dummy); Specialization; Number of fields covered.
Revenues	Total revenues of the university; Tuition and fees; Government appropriations; EU and other international funding; Private funding (profit and non-profit); Asset revenues; Other revenues.
Expenditure	Total expenditure; Personnel expenditure, if possible divided between personnel categories; Current expenditure; Capital expenditure; Other expenditure.
Personnel	Total academic staff (Headcount or FTE); Full professors; Associate professors; Researchers; Other academic staff; Technical and administrative staff.
Educational activities	Number of enrolled students; Number of foreign students; Number of graduates (when applicable divided in long cycle and short cycle graduates); Number of PhD students; Number of PhD degrees; Number of master students; Number of master degrees
Research and technology production	ISI publications; Patents; Spin-off companies; R&D revenues; R&D expenditure

Source: AQUAMETH, in Bonaccorsi, Daraio, Lepori and Slipersaeter (2007b)

The design of the strategy and choice of variables for the full data collection requires also taking into account the results of the mapping exercise of higher education statistics in the European countries performed in the first phase of EUMIDA. This exercise displays, first, large problems of data availability and reasons for non-availability in many countries, but also wide differences between types of data, with the better situation found for data on students and degrees and the most difficult one for research output. Moreover, the design strategy needs to consider also issues of feasibility and required investment, since the number of variables and of HEIs is likely to be very substantial.

This means that, while for the core set of variables a unique strategy was followed – where in principle all data should be collected for all institutions in all countries – for the full set of data a *variable geometry approach* applies:

- First, the level of detail in data takes into account also availability and effort required for data collection; hence, more data will be collected concerning students and degrees – where most information is routinely collected in the UOE data collection -, while much less detail is required for expenditure and finances;
- Second, due the national particularities, some limitations of the perimeter of institutions are accepted against the broader perimeter of research-active institutions;
- Third, it has to be envisaged that the perimeter of available data is different for some countries, due to problems of availability or reasons for non-availability of some types of data;

- Fourth, a broader set of sources than data collected by national statistical authorities has to be envisaged in some cases, even if we acknowledge the organisational and quality problems that might come from the use of non-statistical data sources;
- Finally, concerning research and third-mission output results of the mapping exercise shows that, with the exception of data on PhD degrees, it can be hardly envisaged to get meaningful data at this stage. Therefore EUMIDA will instead realise a number of methodological experiments to prepare future data collection.

2.2.4 Existing framework of UOE data collection and R&D statistics

It is important to remind that two already existing data collections (and methodological manuals) are closely related to the EUMIDA data collection and to a large extent provide the foundations for the definitions presented here. We shortly present them in this Section, by outlining their contribution to EUMIDA, as well as specific differences:

- UNESCO-UIS/OECD/EUROSTAT data collection on education statistics (UOE; UOE, 2006) provides internationally comparable data on key aspects of education systems, specifically on the participation and completion of education programmes, as well as the cost and type of resources dedicated to education. Thus, UOE provides for data collection concerning students, degrees, educational personnel, finances and educational expenditure; data is disaggregated by level of education (using the International Standard Classification of Education, ISCED) and by field of education. The UOE manual is largely the reference also for EUMIDA data collection and whenever possible definitions and classifications are adopted; however, at is shall be clear later, there are also differences which are related to the fact that the basic unit in UOE are educational programs, while in EUMIDA higher education institutions. Data collection is managed in most countries by national statistical institutes (NSI), which deliver summary tables with national aggregates to EUROSTAT.
- OECD/EUROSTAT joint data collection on research and development (R&D) statistics provides data on R&D expenditure and R&D personnel, based on the definitions, classifications and procedures for collecting R&D data provided in the OECD Frascati Manual (OECD 2002) and, for EU Member States, following the requirements of Community Regulation CR/753/2004. R&D statistics is based on regular surveys of research performers, even if in some countries higher education R&D expenditure are derived from higher education statistics. The Frascati manual is relevant when looking to research expenditure of higher education institutions; however, its focus is to provide national aggregates of R&D expenditure and this requires to split research and education in HEIs by using staff time as criterion, while EUMIDA considers the whole of activities an individual HEI.

The strategy followed in EUMIDA is to rely as far as possible to definitions, methods and data sources from these existing statistics, but to provide adaptations and improvements specifically related to the fact that EUMIDA is dealing with Higher Education Institutions considered as strategic units and not with programs or with a specific activity inside HEIs.

2.3 Defining the perimeter for data collection

2.3.1 Conceptual problems

While in the Middle Age to identify universities in Europe was an easy task and their population was composed just by some dozens of institutions (Rüegg 2004), today the definition of a perimeter – i.e. the choice on which institutions are to be included in higher education – is far from being a simple and obvious decision. Thus, while the core of universities awarding doctorates is composed by less than 1,000 institutions in the 29 ERA countries and reasonably reliable lists of institutions can be compiled from national sources, estimates of the total number of HEIs circulating in Europe are in the range between 4,000 and 6,000, and might include such diverse institutions as institutes of technology, colleges, military schools, professional training schools, with definitions that may differ across countries etc.

While there is some understanding that being part of higher education is related to specific functions – like delivering some kinds of degrees and performing research- functional criteria alone are not enough to identify a perimeter. As shown extensively by population ecology, categories and distinctions between organizational populations are cognitive constructs related to representations of the world that are specific to some audience (Ruef 2000). In our case, the definition of what universities and higher education institutions are is by and large a political decision, endorsed in most countries through a legal act, with relevant practical implications in terms of status, right to award diplomas and access to public resources. As we learn from history, this socially and politically constructed perimeter is continuously evolving, with educational institutions striving to get into higher education and, if possible, to get a university status, and restructuring and mergers of HEIs continuously taking place especially outside the traditional university sector (Kyvik 2004).

For the construction of a European census this raises two problems: first, the picture we get of European higher education and of its institutional diversity will to some extent depend on choices concerning the perimeter; second, there is enough evidence that definitions of higher education are different between countries and this is likely to affect comparability.

The solution adopted in EUMIDA is based on the combination of some basic functional criteria to identify HEIs and of flexibility in their application to take into account national specificities. Thus, we considered that a minimum requirement is to deliver degrees at the tertiary level, i.e. at the levels 5 and/or 6 of the international classification of educational

degrees (ISCED; UOE 2006). This is a well-established and accepted classification of educational programs and much care has been taken in the latest revision of the ISCED classification to improve the delimitation between secondary and tertiary education.

However, ISCED classification refers to educational programs and not to institutions; hence, it includes also small-scale educational activities offered by providers whose main mission is different, like vocational degrees delivered by professional associations in countries like Germany and Switzerland. This raises both conceptual and practical issues. First, these cases do not correspond to a common understanding of what *educational institutions* are, since they are individual programs inside organizations with a non-educational mission. Second, in some countries there is a very large number of these providers, which account for a limited share of students and degrees and for which data availability is very problematic.

Thus, the EUMIDA data collection Handbook defines *higher education institutions* as entities which are recognizable as distinct organizations and whose main activity is providing education at the tertiary level (ISCED 5 and/or 6), as well as R&D. Recognizable means that the perimeter of these institutions can be identified rather unambiguously, they have an internal organizational structure and, at least in principle, their own budget. Further, the Handbook provides a number of examples and a set of delimitation criteria, including the main activity, graduation, recognition as part of the national higher education system, size and visibility (see EUMIDA Handbook, Annex 2+5, for full definitions).

Examples of higher education institutions to be included are universities (PhD awarding), as well as universities of applied sciences (Fachhochschulen, Polytechnics). Other examples are Colleges of Arts and Music, theological schools, schools of pedagogy, distance education universities. Military academies should be included as separate institutions, unless their expenditure are already accounted for in other institutions in the same country.

On the contrary, institutions offering only services for education (but no curricula) are not to be included in the EUMIDA data collection despite their inclusion in the UOE data collection. Research institutions, like public research institutes and Academy of Sciences, whose principal mandate is performing R&D, are also excluded (on the basis of the main activity criterion) even if they are delivering some educational activities.

However, this approach leaves to national statistical institutes the responsibility of deciding how to apply the definition and the criteria in their own country, to take decisions in unclear cases and to deviate for good reasons from the Handbook. Besides taking stock of their knowledge of national systems, this approach is motivated by a political reasoning: a **European statistical system on HEIs** – which has to be built through the cooperation of national statistical institutes – can be viable only if it is accepted at national level and it conforms to national practices in identifying what higher education is.

The perimeter for EUMIDA data collection is by definition smaller than the perimeter of UOE data collection. As we shall discuss more in detail in Chapter 3, the coverage of tertiary education in terms of the number of students is 91% on average in the ERA countries.

2.3.2 Identifying research active institutions

A second issue concerned the identification of so-called *research-active institutions,* for which a richer set of data is collected (extended set of data). This reflects a wider understanding that, differently from educational statistics, a statistical system on European universities should focus on those institutions performing some kind of research activities. At the same time, it is envisaged to go beyond the core of research-intensive universities, as identified for example in international rankings, to cover the broader and more diverse landscape of institutions performing research for different purposes and at different levels.
The complexity of this issue comes from the diffusion of the research mandate beyond PhD-awarding institutions, which makes all kinds of distinction difficult, and form the normative value of the research mandate. As a matter of fact, many HEIs in Europe are striving to get this status (Kyvik and Lepori 2010). Since statistical systems are, in this respect, powerful instruments of institutionalization, actors are likely to react to the definition of research activity in terms of status and political implications.
Statistically, the border is difficult to draw: with the emergence of the research mandate in the non-university sector, using PhD as a criterion is no longer possible and there is good chance that some institutions having the right of awarding PhD in unitary systems are less active in research than non-PhD awarding institutions, particularly in countries like Norway, Finland or Switzerland. Thresholds in terms of volume are also of difficult use, first because there is no natural cut-off point and second because of problematic data quality, e.g. concerning R&D expenditure. The reality is that beyond the top international universities the distribution of research activities is more gradual, with a long tail of institutions performing some (and different types of) research.

The choice made in EUMIDA is to consider the degree of institutionalization of research activities as the key criterion, even if the volume might be small. This is relevant since an institutionalized research mission is likely to drive the strategic choices of the considered institution, but also the representations of its functions from other actors. Thus it can have profound practical implications. Additionally, this definition can be operationalised in terms of criteria which can be verified rather easily, e.g. through the analysis of official documents or websites.
Again, the strategy followed is to provide a general definition explaining the rationale for the category, as well as a number of inclusion and exclusion criteria as a support for choice, but leaving to national authorities the responsibility of taking the final decisions (see the next Table).

Table 4. Definition of research-active institutions

Among the whole population of higher education institutions, we distinguish the research-active ones, i.e. those having an institutionalised research activity. This distinction is relevant because of the specific functions and organizations of these institutions.

The definition of research active does not imply a specific level of *research intensity* and care should be taken in distinguishing between research-active and research-intensive institutions (exceeding some threshold, like the one used in the Carnegie classification). However, it implies that research is considered as constitutive part of institutional activities and is organised institutionally and with a durable perspective. Criteria for inclusion are then the following:

- The existence of an official research mandate.
- The existence of research units institutionally recognised (for example on the institutional website).
- The inclusion in the R&D statistics (availability of R&D expenditure data), as sign of institutionalised research activity.
- Awarding doctorates or other ISCED 6 degrees.
- Consideration of research in institutions strategic objectives and plans.
- Regular funding for research projects either from public agencies or from private companies.

Institutions fulfilling at least three of these criteria should be included.

On the contrary, diffused research activities undertaken by teachers on their own interest are not a sufficient criterion to consider an institution as research-active.

As we discuss later in this Report, national choices concerning the selection of research-active institutions have been very different across countries. Whereas in some countries all HEIs in the perimeter are considered as research active, others have made more restrictive choices.

This means that one should be very careful in avoiding over-interpretation of the provided information, especially when comparing different countries. Excluding an institution from the research-active perimeter most likely means that the volume of research is small and that there is limited recognition of a research mission at national level, but does not exclude that some research activities are present and it is quite possible that their volume is higher than in institutions considered to be research-active in other countries.

2.3.3 Multisite institutions

Even if most higher education institutions are basically one-site, with most of their staff and activities concentrated on a single location, there are in fact a few cases of truly multi-site institutions. These include HEIs with foreign campuses, as well as national institutions created through the merger of pre-existing establishments (e.g. University of London; some Fachhochschulen in Switzerland). There are two issues where the distinction between legal institution and local establishments are relevant:

- An issue of governance and strategic decision-making, where local establishments might just follow strategies decided at the level of the whole institution
- An issue of relationship with the regional environment, where in fact distribution of activities across regions might be more relevant than the location of the main seat; this is relevant, for example, concerning the origin of students.

According to the information gathered, there are relatively few cases where this distinction is relevant, but this issue might become increasingly relevant with the internationalization and Europeanization of higher education. Therefore, the EUMIDA Handbook (see Annex 2+5) includes the option of distinguishing between the main seat - meant the place where the headquarters are located and main decisions at institutional level are taken (this might be different than the legal seat for example registered in the trade register) – and the local establishments where educational and research activities are performed.

This possibility has not been systematically exploited in the pilot EUMIDA data collection, but it is strongly advised to introduce it in the routine data collection, concerning the identification of local establishments – focusing on foreign campuses and establishments located in other NUTS regions -, while it is not suggested to collect institutional-level data for local establishments because this would require significant changes in national data collection practices.

2.3.4 Tracking of demographic events

While many HEIs display a long history of continuity in time – some universities even being among the oldest institutions overall in our countries- a more in-depth historical analysis shows many changes occurred in the composition of higher education systems, including the birth of new institutions, more rarely their closure, but also the merger of institutions and various types of change of status. This aspect has been until now disregarded in educational statistics, since the main goal was to provide aggregates at program or country level, but it is of prime importance when building a statistical system on individual HEIs.

For example, this will be needed to establish continuity in records, e.g. by linking the number of students in year X of one institution to the number of students in year X-1 of the two institutions which were merged into it. Experience shows that especially in systems going through large-scale restructuring – like in most Central and Eastern European Countries - exploitation of time series will be impossible without sound demographic information.

Thus, building on existing practices in the Business Units Register, EUMIDA suggests the inclusion, within the register, of all variables tracking demographic events like birth, death, mergers, split of institutions (see the EUMIDA Handbook for full details), using a unified system of code identification. This approach has not been used in the pilot data collection, since it covers a single year, but will have to be implemented in the regular data collection.

2.4 Characterizing higher education institutions: the core set of data

As already introduced, in EUMIDA two levels of detail have been implemented, namely a core set of data to characterize higher education institutions and an extended set for research-active institutions only, allowing a quantitative characterization on inputs and outputs for the purposes of a richer analysis of HEI activities (for example benchmarking, efficiency analysis, studies of differentiation).

For the whole perimeter (in total 2,457 HEIs in 27 countries, France and Denmark excluded), the main objective, besides building a census, was to provide a set of variables allowing a basic characterization of the European higher education landscape. This meant that completeness of the dataset was a main concern; variables should be chosen whose availability was expected throughout the whole of Europe and where no significant confidentiality problems were expected. These criteria led for example to the exclusion of several financial variables from the dataset.

Thus, while the overall structure from the U-Map project was maintained, most variables have been with data available in national statistical systems – e.g. for research activities only doctorate and status of being research active was included. We consider this an interesting case of complementarity between an experimental project, building foundations through *ad hoc* data collection, and official statistics. The next Table presents the list of dimensions and variables considered.

We notice that this list is strongly biased towards indicators derived from educational statistics and especially based on the number of students and degrees, the type of data belonging to the core of educational statistics and where a strong effort of harmonization has been already done.

Only experience in regular collection and exploitation of this dataset will show to which extent it is feasible to enrich the set of quantitative variables. A second, more severe limitation, is the bias towards education and, to some extent, knowledge production (number of doctorates) and the lack of variables concerning knowledge exchange and regional engagement (with the exception of the region of establishment): while their importance was recognized, no one of the proposed variables – like students from the region, patents, funding from private sources – was likely to be available with a reasonable effort. This is not just a technical issue: if statistical systems shape our representation of reality and have practical effects on actor's behaviour, we consider very risky to systematically neglect some variables on grounds of availability and resources required.

Table 5. List of core set of variables

Dimension	Indicator	Measure/definition
Identifiers	Institutional code	Country code + numeric identifier
	Name of the institution	National language + English translation (if available)
Basic institutional descriptors	Country	Country code (ISO)
	Legal status	Public/private, following UOE manual
	National type	National type of institution (university, college, etc.)
	Foundation year	Year of first foundation
	Current status year	Year when the institutions got the present status
	University hospital	Dummy variable (0/1)
	Total staff	Full Time Equivalents, following UOE manual
Educational activities	Students at ISCED 5 level	Headcounts
	Students at ISCED 6 level	Headcounts
	Subject specialization	Subject domains with students enrolled (ISC fields)
	Distance education institutions	Dummy variable (0/1)
	Highest degree delivered	Diploma/bachelor/master/doctorate
Research activities	Research active institution	Dummy variable (0/1)
	Number of doctorates	Degrees at ISCED 6 level
International attractiveness	International students	Headcounts (ISCED 5)
	International doctoral students	Headcounts (ISCED 6)
Regional engagement	Region of establishment	NUTS code of the region of the main seat
Knowledge exchange	Not available (see 2.4.7 below)	

Source: EUMIDA

We provide in the next Section a short description of the variables for each dimension; for full definitions and methodological guidelines the reader should refer to the EUMIDA methodological Handbook (Annex 2+5).

2.4.1 Identifiers

The database include a unique identifier for the considered HEI and the name of the institutions, both in the national language and in English if a public translation (e.g. on the website) exists. Institutional names are most useful for the purposes of comparative analysis

since in many cases they demarcate different types of institutions at national level (e.g. colleges or military academies).

2.4.2 Basic institutional descriptors

This set of descriptors provides some basic information on the considered HEI in form of dummy variables or simple categories, like country, the legal status of the institutions (based on UOE classification), and the presence of university hospitals (as in the Frascati manual).

Most of these descriptors have been collected for the first time for EUMIDA and are relatively unproblematic. A complex case is however the year of foundation, which is difficult to identify for institutions with a complex history, including for example mergers and changes of status (for example from professional schools to universities of applied sciences). Hence, the choice was made to include a *foundation year* – meaning the year of foundation of the first traceable ancestor – and a *current status year* – where the institution got the present legal status and configuration. Of course, these choices are likely to have a profound impact on the analysis of the demography of HEIs in Europe.

The only quantitative indicator is the number of staff – in full time equivalents - which has been considered as the simplest and better available proxy for institutional size and which should be available from the UOE data collection. This was an outcome of previous work which showed that students to staff ratios are systematically dependent on the subject domain and thus that using students as a proxy of size would introduce a significant bias towards social sciences and humanities (Lepori *et al.* 2010).

2.4.3 Educational activities

The following indicators have been chosen for this dimension:

- The number of students at ISCED 5 level and ISCED 6 level, complying with the definitions of the UOE manual.
- The specialisation in subject domains, by using the fields of education classification of the UOE manual. For purposes of simplification, in the core set only a yes/no variable is introduced meaning the presence or not of a domain in educational activities. Fields which are considered as very small or marginal can be excluded.
- Distance education institution: this is a yes/no variable meant to design fully distance education institutions with almost no students on the campus.
- Highest degree delivered, distinguishing between diplomas (less than three years), bachelor, master and PhD. This classification has been preferred to the one in ISCED given the wide diffusion of the Bologna reform at the European level.

These indicators provide a rather simple but complete view of the main dimensions of educational activities, without particular burden for data collection.

2.4.4 Research activities

As we shall discuss more in detail later (see Chapter 3), the measurement of research activities is one of the most problematic areas of the HEI indicators and thus EUMIDA data is reduced to minimum both in the core and in the extended set of data for feasibility reasons.
The core set includes only two variables, namely the status of research intensive (yes/no, see 2.3.2), and the number of graduations at the doctoral level (ISCED 6 level), which is a widely used indicator of research intensity in PhD awarding institutions and is readily available from UOE data collection.

2.4.5 International attractiveness

International orientation is a very relevant and emerging dimension for most higher education institutions and is very likely to provide relevant input for classification purposes, distinguishing between local and regional oriented institutions and those more open internationally. Most indicators in this area based on number of students and staff from abroad, as well as on international publication. For the core set of data, it is proposed to keep the following indicators, the first one referring to education, the second to research:

- Number of international undergraduate students
- Number of international PhD students

The choice of these indicators takes into account the fact that data on students is usually more reliable and easily available than data on staff and finances in most countries.
The definition used in the UOE Manual is the following:

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

We followed this definition, although we recognize the difference between the notion of international students as based on the citizenship (foreign students) and the notion based on the country of prior education.

2.4.6 Regional engagement

Presence and contribution to the social and economic development in their region is considered as a very important dimension of higher education, especially for the institution in the non-university sector. Therefore, it would be highly relevant to provide a few indicators in this domain already in the core set of data. However, this would need detailed data collection and, secondly, breakdown of data by region and not at the level of whole institution (which raises difficult issues for multi-site institutions).

Therefore, only the code of the regional activities (based on the European NUTS classification[10]) is included. For multi-site institutions, it is advised that codes are provided for the different local establishments, but this is not yet implemented in the pilot data collection.

We emphasize the fact that providing a geographic localization of universities will open the way for a large number of future analyses. For example, data on students or staff might be combined with largely available demographic data to provide indicators of intensity (e.g. number of students per thousand inhabitants, or per population in the relevant age cohort; number of academic staff per thousand inhabitants, or per thousand active population etc.). These indicators would convey different information than the currently available demographic indicators of human capital, which refer to the place of residence.

In addition, some indicators might be developed by field of education, giving a much more fine-grained representation of different kinds of human capital (e.g. number of undergraduate students in engineering per thousand inhabitants, etc.).

2.4.7 Knowledge exchange

Knowledge exchange refers broadly to the transfer of activities to economy, society and culture. We consider important to include indicators on this function since it has become increasingly relevant for higher education institutions and there are large differences between individual institutions in this respect. While the importance of this dimension was recognized, no one of the proposed variables – like students from the region, patents, funding from private sources – was likely to be available with a reasonable effort and thus no such indicator is included in the core set of variables (for a further discussion see Chapter 3 in this Report).

We keep a separate category in order to call the attention of the lack of largely available indicators across all HEIs.

2.5 From characterization to a broader set of variables

Data Collection 2 has involved the development of a set of *quantitative variables* characterizing inputs and outputs of higher education institutions. This represented a more difficult challenge given the well-known problems of data availability and quality concerning, for example, financial variables and research output (Bonaccorsi *et al.* 2007b). Once the feasibility of the development of a census had been demonstrated, the emphasis was shifted towards picking the variables needed for analytical purposes, even if it could be presumed that not all of them would be available for all countries. The underlying reasoning was that by demonstrating needs and showing that a number of countries could deliver data of

[10] See http://ec.europa.eu/eurostat/ramon/nuts/splash_regions.html

acceptable quality, pressure would be put on the European statistical system to evolve towards a more standardized data collection.

Table 6. Variables for the extended set

Category	Variable	N. of variables	Breakdown requested
Expenditure	Total expenditure	4	Current expenditure Personnel expenditure Non-personnel expenditure Capital expenditure
Revenues	Total revenues	3	Core budget Third-party funding Student fees.
Personnel	Number of personnel	10	Academic and non academic personnel. For academic personnel: breakdown national/foreign. For academic personnel breakdown by fields of science.
Educational activities	Enrolled students at ISCED 5 and 6	22	By fields of education Between national and foreign students. By level of education
	Number of graduations at ISCED 5	44	By fields of education Between national and foreign students. By fields of education
	Number of graduations at ISCED 6	11	Between national and foreign students.
Research involvement	R&D expenditure	1	No breakdown requested
	Patents	1	No breakdown requested
	Spin-off companies	1	No breakdown requested
	Private funding	1	No breakdown requested

Source: EUMIDA

Feasibility considerations were taken into account by adopting different levels of detail for each variable – hence more disaggregation is required concerning students and degrees, where most information is routinely collected in the UOE data collection, while much less detail is required for expenditure and finances – and by accepting that some countries further reduce the perimeter for data collection to reduce the workload to an acceptable level. Further, it was envisaged that a broader set of sources could be used, even if we were

aware the organizational and quality problems that might come from the use of non-statistical data sources.

The list of variables presented in Table might look as a modest extension of the core set, but in practice it implies a jump in the number of numerical variables, from 6 to 98. Moreover, disaggregating further data (e.g. for students) clearly makes problems of data quality and of consistency more difficult to handle. When looking to the outcomes of data collection, the picture is as follows. In the following of this Section we provide additional detail on the chosen variables and definitions; for complete methodological information the reader should refer to the EUMIDA methodological Handbook, whereas for information on data availability and emerging quality problems to Chapter 4 and 5 of this Report. We discuss separately data on research and technology outputs in a separate Chapter (3).

2.5.1 Revenues and expenditure

This data is known as rather problematic and difficult to use for international comparisons, because of issues of perimeter (e.g. inclusion of ancillary services), different accounting practices and the lack of sector-specific deflators.

Thus, the choice was made to reduce to the minimum the level of disaggregation; concerning expenditure, the only disaggregation suggested is between personnel, other current expenditure and capital expenditure.

The chosen perimeter corresponds to expenditure inside institutions as defined in the UOE manual (UOE Manual section 2.9), covering all types of goods and services, namely both educational goods and services, R&D and non-instructional services. Thus data collection follows strictly the institution-based approach of EUMIDA. It must be however acknowledged that this criterion can lead to comparability problems since inclusion or exclusion of ancillary services (for example meals or transportation) is variable according to the country considered (UOE Handbook 2.9; see Bonaccorsi *et al.* 2007b).

We also followed the UOE Manual in measuring capital expenditure – thus, accounting capital acquisitions fully in the year of expenditure and not recording depreciation of capital assets as expenditure.[11] While this is usual practice in public accounting, we recognize that this approach is becoming problematic since HEIs in many countries increasingly have

[11] See the following definitions: "In keeping with the system used by many countries to record government expenditure and revenues, the UOE educational expenditure data is compiled on a cash accounting rather than an accrual accounting basis. That is to say that expenditure (both capital and recurrent) is recorded in the year in which the payments occurred. This means in particular that:

- Capital acquisitions are counted fully in the year in which the expenditure occurs;
- Depreciation of capital assets is not recorded as expenditure, though repairs and maintenance expenditure is recorded in the year it occurs;
- Expenditure on student loans is recorded as the gross loan outlays in the year in which the loans are made, without netting-off repayments from existing borrowers.

(…) A consequence of the accounting basis used is that sharp fluctuations in expenditure can occur from year to year owing to the onset or completion of school building projects, which, by their nature, are sporadic." (UOE Manual, vol. 1, section 2.9.2).

private-type accounting practices. We consider this certainly as an area where further improvement is required at the level of statistical systems.

Concerning revenues, a disaggregation is required between core budget, student fees and third party funding, defined as funding earmarked to specific activities and institutional units, in most cases also limited in time, both for private and public sources (Lepori *et al.* 2007; Jongbloed 2008).

Third-party funding includes specifically grants from national and international funding agencies for research activities, contracts from public bodies and private companies for specific research and services, fees from companies for educational services, donations and direct support target specific activities (e.g. chairs).

Following the literature in the field, we consider that this classification is more relevant for analyzing HEIs operations than the one provided by the UOE Manual between public and private sources and that overall this distinction is possible for most of the HEIs revenues (see the Handbook for definitions).

2.5.2 Personnel

This data provide information on the staff employed by higher education institutions conforming to UOE Manual definitions and practices. This data is extremely relevant since they provide much information on resources available to institutions; also, cross-country comparisons are considered to be more reliable when using personnel data than financial data.

Breakdown of staff by categories is highly relevant for detailed analysis; however, it is also likely to significantly affect the burden for data collection. Following breakdown is required in the EUMIDA data collection:

- Personnel categories: breakdown between academic staff and non-academic staff. This distinction is highly relevant given the diversity of functions between the two categories and should be readily available since it is already foreseen in the UOE data collection.
- Field: breakdown of academic staff by field of activity using the Field of Science classification of the Frascati manual. FOS is preferred since it is a better classification for HEI organizational units than fields of education and since R&D personnel data is collected using FOS.
- Nationality: breakdown of academic staff between national and foreign staff using nationality as a criterion. This breakdown is relevant to analyse internationalization of HEIs.

All data should be in Full Time Equivalents. This choice is important since at least outside universities a large number of staff is likely to be employed only part-time and thus use of headcounts is likely to strongly affect comparisons.

PhD students are a major issue concerning university staff, as it was shown in the AQUAMETH project (Bonaccorsi *et al.* 2007a). As a general rule, PhD students should be considered as an input for higher education institutions since in fact most of their time is devoted to producing outputs of the institution itself, in form of dissertations, scientific publications, support in research, administrative duties. However, there are some indications that inclusion of PhD students in personnel data is highly variable depending on the country; firstly, in some countries PhD students are financed by national grants and thus they might not be included in personnel statistics, while in other countries their level of employment might vary depending if the time devoted to the dissertation is included in the working contract.

Given the large number of PhD students in some countries, their different status is likely to raise relevant comparability issues.

EUMIDA cannot provide a solution to this issue; however, it is required to include a specific metadata in the data collection specifying the employment status of PhD students.

2.5.3 Educational activities

This data represents the core of educational statistics and this is explains why much detail was required (77 variables are in this domain). Data cover both enrolments and graduations and require disaggregation by field of education, by program level – adopting the Bologna classification in bachelor, master and doctorate – and finally by origin (country of prior education if available, nationality otherwise).

There are relatively few methodological issues in this area, the most important one being probably is differences in counting graduations between programs with 1st and 2nd level qualifications (e.g. bachelor and master) and those with single-cycle 2nd level qualifications without a 1st level qualification (e.g. the old licence system).

The approach recommended by the UOE manual to count students which graduate for the first time at 2nd level degrees also in 1st level degrees might be difficult to apply at the institutional level (UOE manual 3.3.4). The approach proposed in EUMIDA is to single out those 2nd level degrees (e.g. masters) without a 1st level qualification in a separate line; the feasibility of this approach needs to be checked on the basis of national information (see Chapter 3 of this Report).

3. Measurement of research activities and outputs

3.1 Conceptual and methodological problems

Research is considered as a central activity of Higher Education Institutions (HEIs), the one being the most noble and valuable, giving them their specific character as academic institutions. Also, the rationale for launching a large-scale project on HEI statistics was by and large to better seize their contribution to the development of the European Research Area. Consequently, strong priority was given to the measurement of research output.
This focus should not be interpreted as a statement that research is the main output of universities. In particular, the important role of universities in providing a highly skilled staff to European businesses (including SMEs) should not be neglected. In the context of a renovated interest on intellectual workforce in the labour market, this function remains extremely important: in some countries, future availability of young experts may be limited by demographic developments towards an ageing society, while at the same time, the need for staff specialisation is driven by the increasing relevance of complex technologies for the European competitiveness. Therefore, research is one important activity of HEIs but not the only one.

A closer look at the HEI landscape reveals that their activities are not uniform. Rather, the different HEIs have different missions and - as a consequence - different profiles. This diversity is the starting point for the so-called U-Map project where the different dimensions of HEIs' activities are analysed, using the example of 100 European HEIs.[12] On their website, the authors of the project describe the background and context of their work in the following way.

> "*The concept of diversity has moved rapidly up the political agenda of European higher education over the past decade. The development of the European Higher Education Area (EHEA) and the European Research area (ERA) has clearly contributed to this. Global debates about world class universities and international competition in higher education and the growing popularity of rankings and league tables have both triggered awareness that the diversity of European higher education is a potential strength but that the better understanding of this diversity is needed*." (http://www.u-map.eu/about.doc)

The basic concept of U-Map postulates that HEIs have different profiles and that only HEIs with similar profiles can be compared in a meaningful way. The authors highlight the following dimensions of activity:

[12] See Van Vught et al. (2010) and http://www.u-map.eu/

1) Teaching and learning
2) Student profile
3) Research involvement
4) Involvement in knowledge exchange
5) International orientation
6) Regional engagement

Thus within such a broader concept, the data on research output exclusively refer to the third dimension. However with the EUMIDA data presents a much broader data set than is available in U-Map, so that the variation between profiles can be analysed in a more detailed way.

An important methodological issue of U-Map is the concept to explicitly show the different the different dimensions of activity and to present profiles, rather than to merge them to one final performance value by means of composite indicators. In a similar way, the different data collected in the EUMIDA data set should be considered as indicators of different activity dimensions and not combined in a simplistic way in order to achieve a sufficient transparency. In this perspective the EUMIDA work can be considered as a feasibility study as to the broader collection of data on HEIs for different activity dimensions.

However, this is an area of where few commonly accepted standards exists for the production of data at the institutional level (with the exception of the number of doctoral degrees); there are some recognized indicators of outputs – like scientific publications or patents, but none of them has reached the level of standardization and of acceptance for its systematic production on a broad set of European HEIs. Moreover, in most countries national statistical institutes do not produce indicators on research output. Thus, they do not fit easily into the institutional framework of the EUMIDA project.

Thus, for example, bibliometric indicators based on international databases – Web of Science, Scopus – are widely accepted as measures of international research reputation, but their validity in the fields of social sciences and humanities is questionable, and are usable only for the few hundred research intensive universities (van Raan 2007). At national level, the EUMIDA data collection has shown that different data sources from national evaluations exist, but these are not comparable and in most cases rest on different definition of what a scientific publication really is.

The role of EUMIDA in this area was not to provide solutions ready to be integrated in the statistical system, but to highlight a few areas where currently experiments are being developed and to promote awareness that in future some of them might generate widely accepted indicators, for which statistical procedures will have to be established. Moreover, having a census of HEIs in Europe will help to better understand the value of existing data on research output and might help to solve attribution problems by more easily identifying HEIs and their perimeter.

To this purpose, a few experimental data were requested on a separate sheet, marking them clearly as option, to check if some of them could be included in the regular data collection.
In this Chapter we will:

- Review the research activities and outputs for which standardized measures are available;
- Discuss future feasibility work to develop standardized measures for areas currently not covered;
- Provide a pilot feasibility study on the utilization of the Scopus dataset on universities not covered by the Leiden ranking of top 200 European universities.

3.2 Standardized measures

In addition to indicators already discussed in Chapter 2 with respect to various dimensions of diversity, there are others specifically related to research output. In this Section we will discuss the main methodological problems, while in Section 3.3 a preliminary presentation of results from Data Collection 2 will be introduced.

3.2.1 R&D expenditure

Concerning *R&D expenditure* there is a well-defined procedure to produce this data based on timesheets of academic personnel (OECD 2002). However, there are serious issues concerning these measures. First of all, it is questionable to which extent education and research are separable activities. Second, it is not at all clear whether academic personnel are able to report correctly their time investment in each activity (Lepori 2006). Even disregarding these conceptual issues, many European countries do not follow the procedures recommended in the Frascati Manual (OECD 2000). A detailed analysis of the OECD Sources and Methods database and of a recent EUROSTAT questionnaire shows that few countries perform a survey of the use of time of HEI personnel, either each year or regularly every 3-5 years (using coefficients to interpolate values. This data can reasonably used at the institutional level.
However, in many countries, including Italy, France Germany, calculation of R&D expenditure is based on national coefficients, sometimes derived from older surveys or even created following "rules of thumb" by national statistical institutes. These datasets are clearly not usable at the institutional level since they are based on national-level coefficients. It would be meaningless to allocate nation-level average coefficients to individual units.
As a matter of fact, in the EUMIDA pilot, data on R&D expenditure has been requested in order to check for their availability and methodological problems (see Chapter 7 of this Report).

3.2.2 Funding from the private sector

Funding from the private sector is a rather widely used indicator of knowledge transfer to the private economy since a large share of private funding of universities is linked to research contracts, usually from industry. In principle, this data should be readily available from university accounts and regularly collected in official statistics both for total private funding and for funding for R&D and aggregate figures at national level are provided, e.g. by the OECD (2008).

In practice, experience from previous projects showed that there are significant methodological problems with this data. First, private funding covers different situations, with different meanings in terms of public-private relationships, like unrestricted grants (such as from private foundations and trusts), rent of facilities and/or infrastructure, sales of services, sales of gadgets, or contract research from private organizations. Second, private funding is particularly affected by issues of perimeter concerning ancillary services, so that the delimitation between private and non-profit is not always easy. Third, private contributions cover a small share of total budget of HEIs and are thus particularly affected by data quality problems.

In the EUMIDA pilot, data has been requested on both total funding from private companies – thus excluding student's fees – and on funding for R&D activities (see Chapter 7 for a discussion of the results).

3.2.3 Patents

For describing technology production and technology transfer, as part of third mission, patents seem to be one of the most attractive indicators. Their specific advantage is the availability in various databases and the possibility of a breakdown by technology fields (patent classes), date of application, inventor, assignee etc. However, for European universities empirical research showed that in many cases where a patent application originates from a university the assignee is not a university and the name of the inventors do not report the academic affiliation. Furthermore, the legal situation as to university patents is quite different in member countries, so that counts of patents based on institutional affiliations, as specified in the patent application, cannot be easily compared.

An alternative approach based on names of inventors – starting from a list of names of academic staff- has been developed in the Italian case (Balconi *et al.* 2004) and then successfully extended to France, Netherlands, Sweden and UK (KEINS database; Lissoni *et al.* 2008). Using this methodology, it was discovered that the vast majority of patents invented by academicians are assigned to subjects different from the university institution. This methodology proves to be very precise, but very time consuming. Presently new approaches are tested where inventor names for applications in specific years are matched

with author names of publication databases. Positive experiences with this methodology were already made by Noyons *et al.* (2003a & 2003b), but limited to specific areas.
In the EUMIDA pilot, data based on both methodologies (institutional affiliations and names of inventors) have been requested.

3.2.4 Spin-off companies

Research carried out within the PRIME Network of Excellence, under the project Rebaspinoff coordinated by Philip Mustar, has provided a preliminary effort to collect and standardize data on spinoff companies created by European universities (Mustar *et al.* 2006; Mustar, Wright and Clarysse 2008; Clarysse *et al.* 2007; Wright *et al.* 2008).
The research showed a much complex situation than anticipated concerning the notion of spin-off, the perimeter and the type of relationships with the university. Common themes in relation to the classification of spin-off companies are (1) spin-off creation and (2) spin-off development. The dimensions that differentiate between firms are the type of resources, the business model and the institutional link.
In EUMIDA, it was suggested as a manageable solution to consider only spin-off company which are recognized in the official documentation of a university as such, following a formal internal procedure, and this resulting in the website of the university, or alternatively any written official documentation, mentioning explicitly and listing by name all recognized spinoff companies.

3.3 Output of research activity: preliminary evidence from the EUMIDA dataset

In this Section, the major results with regard to the output of research activity at European universities as reflected in the data collections of EUMIDA are presented. After a brief general reflection, the main output indicators of the Data Collections 1 and 2 are packaged. Then the possibilities of including potential additional output indicators are discussed.
The goal of this Section is *not* to provide a general description of indicators, which is offered in great detail in Chapter 7, together with breakdown of data by country and category. Rather it aims at discussing potential and limitations of available indicators for describing the research output.

3.3.1 Students and graduates at the ISCED 6 level

In terms of activity indicators the number of students at the ISCED 5 level is an indicator in the performance dimension of teaching and learning, differentiated by field of education for the student profile – thus for the dimensions 1 and 2 according to the categories of U-Map. Students at the ISCED 6 level and doctorates awarded are major elements of the research involvement (dimension 3 of U-Map) and should be interpreted as research output

indicators. A HEI offering doctorates as the highest degree should have a relevant research activity for providing an appropriate supervision and support for doctoral students. Furthermore the preparation of a doctoral thesis is generally linked to research. So students at the ISCED 6 level may be considered as intermediate output indicators for the provision of young research academics. At the same time they are a direct indicator for research input, as the majority of ISCED 6 students is active in research. The great advantage of the collection of microdata for all universities is that both perspectives (PhD students and graduates as inputs and/or outputs) can be pursued in empirical work.

In the core data set of EUMIDA (Data Collection 1) 2,457 HEIs are included, whereof for 2,333 HEIs (equivalent to 95%) data for student numbers are provided. In total 15,719,854 students at the ISCED 5 level are covered. Furthermore 6 HEIs display exclusively students at the ISCED 6 level, i.e. are graduate school without undergraduate courses. It is realistic to assume that the HEIs displaying neither students at the ISCED 5 nor ISCED 6 level are less relevant in terms of student number, and that the coverage of students at the ISCED 6 level is complete.

At the ISCED 6 level, the core data set covers 531,370 students and 92,631 doctorate degrees awarded. The number of institutions offering a doctorate as the highest degree is 885 equivalent to 36% of all HEIs. Further 5 HEIs report offering an intermediary ISCED 6 qualification and thus have students at the ISCED 6 level. In total 890 HEIs have students at the ISCED 6 level. Thereof 850 HEIs report to be research-active equivalent to 96% of all HEIs with ISCED 6 students. The remainder comprises art colleges, theological academies, defence universities, or specialised HEIs in management or finance. However the general assumption that the provision of doctorates as highest degree and the existence of students at the ISCED 6 level is linked to research is supported and thus the use of student data on the ISCED 6 level as one indicator of research output.

Nevertheless, it has to be taken into account that a broader definition of research-active HEIs was adopted for EUMIDA. Criteria for inclusion were the following:

- The existence of an official research mandate,
- The existence of research units institutionally recognised (for example on the institutional website),
- The inclusion in the R&D statistics (availability of R&D expenditure data), as sign of institutionalised research activity,
- Awarding doctorates or other ISCED 6 degrees,
- Consideration of research in institutions strategic objectives and plans and
- Regular funding for research projects either from public agencies or from private companies" (EUMIDA Consortium 2010).

In consequence more research-active HEIs are displayed in the data set. Therefore, 555 HEIs of the 1,566 HEIs without ISCED 6 students report to be research-active, thus about 35%. However, the results of the bibliometric analysis (see Section 3.4 below) provide some

evidence that the research output of these HEIs without doctorate is substantially lower than that of HEIs with doctorate.

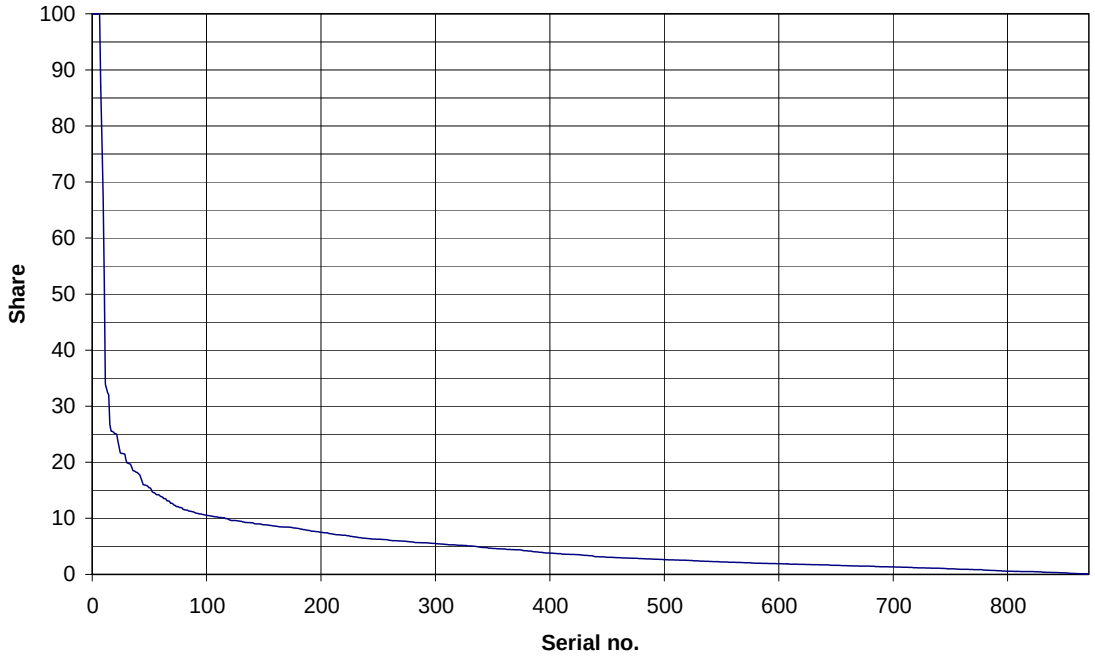


Figure 2. Share of ISCED 6 students within all students of a HEI (in %) for 870 European HEIs

Source: EUMIDA

An interesting indicator to describe the research orientation of HEIs is the share of ISCED 6 students within all students of a HEI. In the data set 870 HEIs provide data for students at the ISCED 6 level. Therein the maximum share of ISCED 6 students within all students is 100% the minimum 0.1%, the median 3.3%. The distribution of this share is extremely skewed, as shown in Figure 2. The HEIs with very high shares of ISCED 6 students are generally quite small (in terms of student number) and specialised on specific fields such as theology, arts, or specific technologies. The HEIs with very low shares are primarily teaching/education-oriented. The intersection with HEIs without ISCED 6 students appears to be diffuse. The "standard" universities can be found in the range of shares between 2 and 8%; then this share can be interpreted in terms of research orientation. Thus, this indicator has to be interpreted in context with other information such as size, specialisation, mission, legal status etc. This broad variety supports the introductory reflections on different HEI profiles and missions.

Further analysis of ISCED 6 students can be realised on the basis of the extended data set collected in the project. This set primarily comprises HEIs labelled as research-active[13]. All in all, the extended set comprises 1,405 HEIs (1,362 with complete data), therein 846 HEIs

[13] The detailed definition for the extended data set can be found in Chapter 2 of this Report and the main findings are reported in Chapter 7.

with at least one ISCED 6 student. The extended set comprises 13.584 million students at the ISCED 5 level and 542,261 students at the ISCED 6, i.e. 91% and respectively 100% of the core set.

In the extended set the differentiation by field of education is included as shown in Table 7. This differentiation is displayed for 642 HEIs. The data in Table 7 shows that the largest share of students at the ISCED 6 level are associated with the Natural sciences, but also the shares in the Social sciences, Humanities, Engineering and Health are substantial. The relation of ISCED 6 to ISCED 5 students reveals that particularly in the Natural sciences and Agriculture the share of ISCED 6 students is above average. Comparing the distribution of ISCED 5 students in all HEIs to those in HEIs with ISCED 6 students, the latter have a stronger relevance in Humanities, but in general the distributions are similar.

When number of ISCED 6 students is used as an indicator of research activity, it takes particular relevance for Education, Humanities and Social sciences, as these fields are not appropriately represented in bibliometric analyses. Thus, the number of ISCED 6 students is unique for assessing the research orientation in these fields. Furthermore this indicator proves to be important for smaller HEIs, as their publication activity recorded in bibliometric databases proves to be quite low and erratic – also in the Natural and Medical sciences.

Table 7. Share of students by field of education and relation of shares of ISCED 6 to ISCED 5 students

No		General	Education	Humanities	Social sciences	Natural sciences	Engineering	Agriculture	Health	Services
1	ISCED 5 (all HEIs)	0,04	7,04	13,22	36,36	11,02	16,49	1,87	11,19	2,77
2	ISCED 5 (HEIs with ISCED 6)	0,05	7,27	14,49	35,41	11,89	15,45	1,52	11,71	2,21
3	ISCED 6	0,11	3,60	13,25	17,71	18,75	12,46	1,88	11,13	0,93
4	Relation 3 : 1	2,56	0,51	1,00	0,49	1,70	0,76	1,00	1,00	0,34
5	Relation 3 : 2	2,08	0,50	0,91	0,50	1,58	0,81	1,23	0,95	0,42

Source: EUMIDA

3.3.2 Internationalization

Internationalization is a specific aspect of the output of HEIs and represents a specific dimension (dimension 5) in the U-Map project. With regard to the output of research, it can be analysed by international co-publications in bibliometric analyses[14] and the share of international students at the ISCED 6 level. Looking at the distribution of the share of international ISCED 6 students within all ISCED 6 students, available for 843 HEIs, the range is between 0% and 100% (Figure 3). Very high shares can be observed for small HEIs and for larger HEIs in Switzerland and the United Kingdom. Thus there are substantial

[14] This aspect is not discussed in further detail in the feasibility study below. However, once the identification of a HEI in a bibliometric database is realised, it could be easily included.

differences between countries, and appropriate comparisons should be made between HEIs of the same country.

A further indicator for internationalisation is the share of foreign academic staff with reference to the total academic staff. However, the academic staff is active in research as well as teaching, so that it is not appropriate to reflect the output of research. In some cases, the orientation of a HEI on international research, indicated by foreign ISCED 6 students, is also visible in the share of foreign staff. But all in all, no correlation can be found.

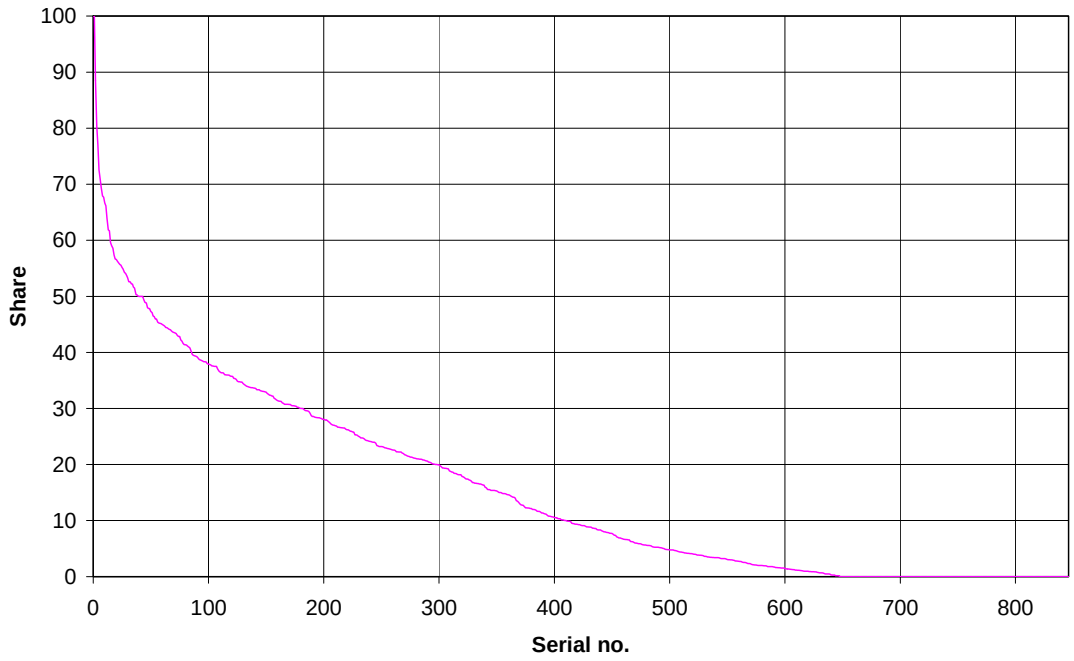


Figure 3. Share of international ISCED 6 students within all ISCED 6 students of a HEI (in %) for 843 European HEIs

Source: EUMIDA

3.3.3 R&D funding

In principle, the funding of R&D is an input indicator for R&D. Nevertheless, some information as to a HEI, the general R&D orientation and the specific orientation on private partners can be derived. Furthermore a calculation of input-output ratios can be performed, if appropriate output indicators are available, and thus the efficiency of the R&D activities may be assessed. Therefore reliable information on the R&D funding is very useful for getting more comprehensive information on R&D.

In the extended data set, 504 out of the 1,322 HEIs of the enlarged data set provide information on R&D expenditure, thus only 38% of the HEIs. The HEIs with data on R&D funding comprise 51% of ISCED 6 students; also in this regard the information is

insufficient. This low share can be primarily explained by the fact that in many countries the HEIs get a lump sum funding without differentiation between teaching, research and other activities. Thus, as discussed at length in Section 3.2.1 of this Chapter, an estimate of the R&D activities may be possible at the country level, but not for individual HEIs.

A technical remark is in order. Data is provided in national currency, and a substantial share of the countries covered in the dataset is not yet part of the Eurozone. This is not a major shortcoming. The data should not be converted based on the official exchange rates, as the latter depend on a variety of factors, which are not useful for comparing the size of R&D activities. Rather, following the current Eurostat practice for R&D expenditure, data should be available in National currency, in Euro, in Purchasing Power Standard and in Purchasing Power Standard at 2000 prices. This remark applies not only to research expenditure, but to all data about expenditure and funding.

In all of the 504 cases with funding data, the data on R&D funding are reported to be lower than the total funding of the HEIs. However, the ratio between R&D funding and total HEI funding has to be interpreted with care, as the total funding comprises the categories personnel, non-personnel and capital funding, where the relations vary largely between HEIs. It seems to be more realistic for the comparison of HEIs to refer the R&D expenditure to the funding of personnel, although a certain share of the R&D funding is spent for non-personnel costs.

A further check of data quality can be performed by calculating the ratio of private R&D funding to total R&D expenditure. In 9 of the 504 cases, equivalent to 1.8%, the figures for the private funding are higher than those for the total R&D expenditure, thus are obviously inconsistent. For the other cases, a distribution according to Figure 4 is found. The share of private funding largely differs by HEI. The cases with very high shares cannot always be associated with small, specialised HEIs, thus cannot be interpreted as statistical outliers. Rather various large HEIs have high shares of private R&D funding. To understand the differences in private R&D funding, a more detailed analysis of the frame conditions of the HEIs would be necessary. In any case, it appears to be not appropriate to interpret high shares of private R&D funding in terms of high performance.

In 10 of the 504 cases the provided figures for private R&D funding are higher than the total private funding, which are again cases of inconsistency.

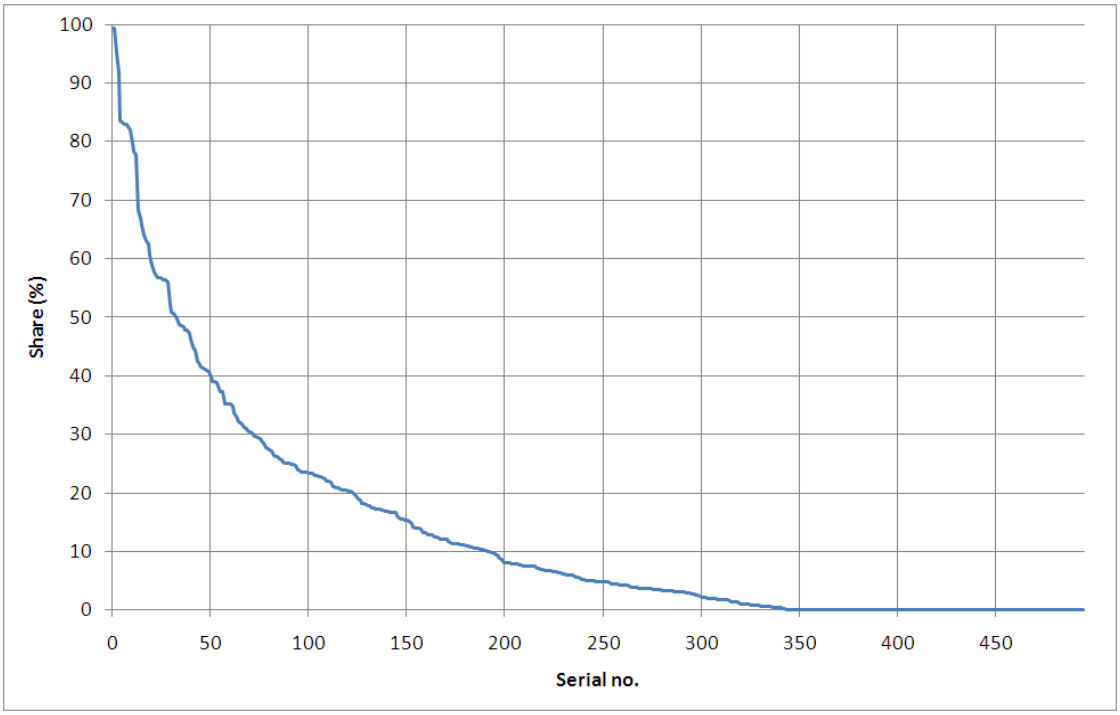


Figure 4. Share of private funding within the total R&D expenditure for 504 European HEIs

Source: EUMIDA

To sum up, the data on R&D expenditure and funding is very useful, if interpreted in the context of other information on the HEIs. However, the share of HEIs providing this data is modest and further efforts are necessary to enlarge the dataset.

3.3.4 Academic patents

Patents from HEIs are considered as important indicators of technology transfer and output of applied research. In any case, they reflect the interest of a researcher at a HEI to transfer his/her scientific knowledge to a concrete technological application. A patent application of a HEI is only useful, if the HEI intends to enter into cooperation with a private partner, as the HEIs have not their own production facilities. In this context, substantial efforts have been made to improve the understanding and the collection of HEI patent data.[15] To describe technology production and technological third mission, patents seem to be one of the most

[15] See for instance OECD (2003). Another important contribution has been recently produced for Eurostat by experts at Katholieke University of Leuven (KUL– B), who developed data production methods for harmonized patent statistics aimed in particular to applicants' sector allocation. This method on assignee sector allocation has been reviewed and improved recently, taking also into account the differences in the role of legislation on the patenting activity in the higher education sector. See the Working document entitled 'Evolution of innovation actors and the influence of legislation'.

attractive indicators. Their specific advantage is their availability from various databases and the possibility of a breakdown by fields, time of application, applicant etc.
However, the collection of this data proves to be quite complex, as the HEI/university does not always appear as applicant on the patent document. Higher education institutions may deposit patents on inventions created by staff affiliated to them. In European countries, the most diffused legal treatment of this situation is based on the principle that patents invented by employees of an organization are assigned to the organization, not to the individual. Noteworthy exceptions are the so called "professor privilege", which was adopted in German speaking countries. In this case the intellectual property was considered part of the endowment of a professor's chair and the economic benefits accruing from patents could be awarded to individual professors. A patent originating in a university can have three types of applicants:

1) The university/HEI
2) An enterprise
3) A private person, in general the researcher (professor, postdoc researcher, doctoral student etc.).

In the cases 2 and 3 the relation to the university/HEI cannot be derived form the patent document, as the university staff appears in case 2 as inventors with their private address or in case 3 with their private address as well (with few exceptions).[16] The share of the cases 2 and 3 can be substantial and surpass the level of 50%.[17] This high share of applications without the HEIs as applicants may be linked to the details of the national regulation. The share of applications with the HEI as applicant is higher in countries with a long tradition of university patent ownership; but the share of other applicants is generally still high.[18] In any case, the assessment of research output exclusively on the basis of patents with universities/HEIs as applicants is misleading. Against this background the following categories for patent applications were enquired in the questionnaire for the enlarged data set:

- Patent applications first applied in the year of analysis (priority year) for which the considered HEI is filed as one of the assignees; and
- Patent applications for which at least one of the inventors is an employee of the considered HEI.

The second question refers to the cases 2 and 3 mentioned above. Out of the 1,322 HEIs of the enlarged data set, only 399 report on their patent activities, whereof 195 display at least one application. These latter are primarily from universities in the United Kingdom, Spain, Italy, Finland and Norway. All these statements refer to the group of applications with the

[16] For more details see EUMIDA consortium (2010).
[17] See for instance Schmoch (2004).
[18] For the USA see Audretsch *et al.* (2006).

HEI as one of the applicants. Only in 4 cases, are data provided for applications with at least one inventor from a HEI. This finding shows that, even if the registration of HEI patent applications is organised in a more systematic way, the information on patent applications with the HEI as origin, but without the HEI as applicant is generally not available. To conclude it proves to be impossible to collect systematic and complete data on patents with a HEI as origin by means of a questionnaire based on data at statistical offices or other public agencies.

The most realistic approach to collect data on patent applications with HEI origin is to search for patent applications with HEIs/universities as applicants and in addition to match data on the names of academic staff with the names of inventors. This method is followed in the so-called KEINS project where the staff lists of academic institutions are matched to the inventor lists of the country examined (Lissoni *et al.* 2008 & 2009). Starting from a methodology developed by Lissoni and colleagues from different countries, the validity of this approach has been demonstrated by extending the analysis to France, the UK, Sweden, Denmark and the Netherlands. In all countries a somewhat similar situation emerges: formal academic patents are only a tiny proportion of all patents invented by academic staff.

The methodology goes as follows:

- Identify the list of names of academic staff from official sources (e.g. ministry of research databases or websites);
- Locate all patents invented by inventors with the same name;
- Cluster all patents with the same name of inventor(s) and check about their disciplinary and/or technological consistency by examining patent classes
- If names are unambiguous and patents fall in the same patent class or in clustered patent classes, then conclude that they are academic patents
- If there are ambiguities in names (e.g. different spellings for the same name, different first name, etc.) check for synonyms by looking at patent classes, trying to allocate unambiguously patents to inventors;
- If the ambiguity persists, control name by name through a web search and, possibly, through telephone interviews.

Although the KEINS approach proves to be very helpful in identifying academic patents not applied for HEIs, the practical application in large data sets, in particular for all member countries, may be difficult. It will be necessary to provide lists of academic staff names for all universities in the EU in regular sequences, as academic researches are quite mobile and staff lists change frequently. Furthermore the lists of academic staff of the HEIs cover only individuals with official positions. While doctoral students or postdoc researchers without formal positions are often not included in the official list, they are however potential inventors.

To cope with these problems new approaches being presently tested, where inventor names of applications in specific years are matched with author names of publication databases in

referring years. If an inventor is identified as the author of a scientific paper, the database contains information about the author's affiliation. Thus the match of inventors and authors helps to indentify the scientific institutional link of inventors. If the institutional link is a university, the origin of a patent application is academic in most cases. The major problem of this approach is that in the cases of frequent names - like Peter Müller in the German case - incorrect affiliations to universities may be identified. A match between technical and scientific fields can reduce this problem, so that, for instance, an invention in electrical engineering is not attributed to a professor in sociology. Positive experiences with this methodology were already made by Noyons *et al.* (2003a & 2003b), but limited to specific areas. Further improvement can be achieved by a regional match between the addresses of the inventors and of the institutions of the authors. In this context, the publication database Scopus proves useful, as it records the full first names of the authors, not only the first letter, so that the precision of match is improved. Furthermore, Scopus has a much broader coverage of the engineering fields than WoS, and many academic patents refer to engineering.

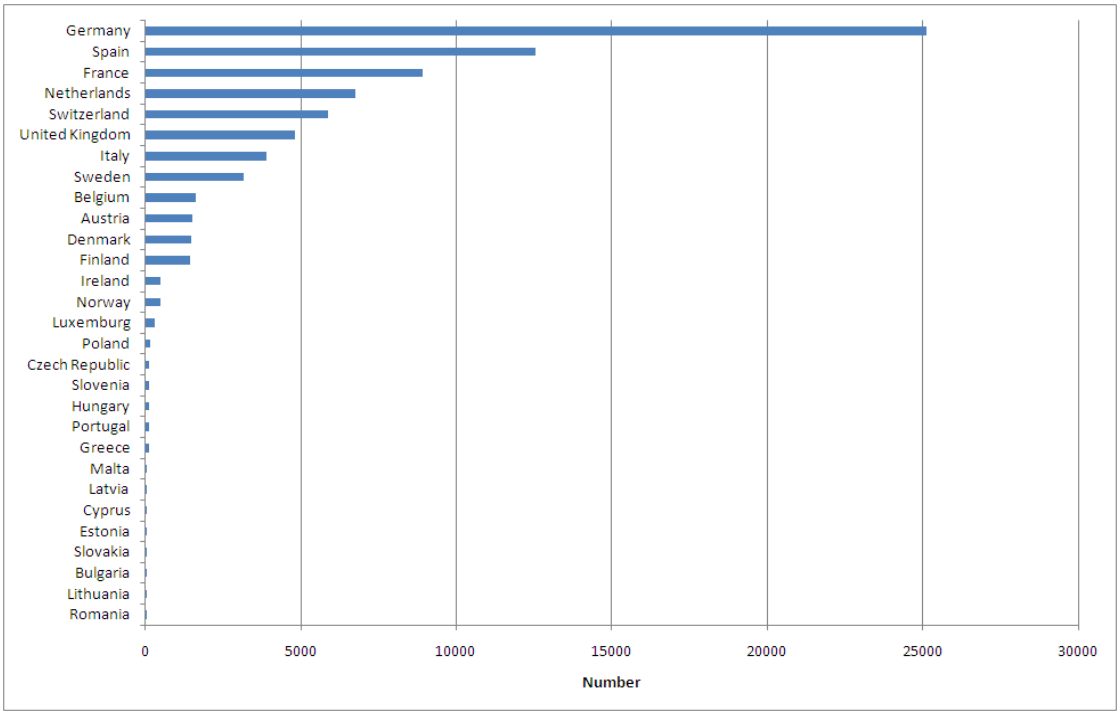


Figure 5. Total number of patent applications at the EPO – including Euro PCT applications – by country in the year 2009
Source: EUMIDA

At present a feasibility study on the match of authors and inventors is being conducted at Fraunhofer ISI (Karlsruhe, Germany) with encouraging intermediate results. The final report will be available in spring 2011 with results for at least 5 European countries. The problem for larger data sets is that the findings for one country cannot be easily transferred to

another one, as the structure and peculiarities of names differ country by country. Therefore a complete coverage of all EU countries appears to be quite ambitious at first sight. However, in many countries, the number of patent applications is quite small in general and it will be easy to identify the few academic patents in this data set (see Figure 5).

3.3.5 Spin-off companies

In addition to patent applications, the number of spin-off companies out of HEIs is considered a good indicator for the transfer orientation of a HEI. In the EUMIDA data set, only 282 HEIs report on spin-off companies thereof only 105 with at least one company. So the reporting is even weaker than in the case of patent applications. If the data is available, it is too low for being used for statistical assessment. Alternative approaches for collecting such data were not suggested.

Based on the results of the EUMIDA collection we do not recommend a further collection of patent and spin-off data in a survey. In the case of academic patents we would rather suggest further developing the matching of inventors and authors to facilitate their identification.

3.4 Additional research output indicators

3.4.1 Publications

3.4.1.1 Introduction

An appropriate description of the activity of higher education institutions has to be based on a variety of different inputs and outputs; but it proves difficult to capture the different dimensions of output by quantitative indicators (Bonaccorsi *et al.* 2007b: 68).

Notwithstanding many legitimate criticisms, publications stand as a lead indicator for the output of scientific research. A more detailed examination of this indicator is necessary (Schmoch and Schubert 2009). However, the number of publications – and even more the citations referring to them – is generally not available at the statistical offices. In order to assess the possibilities and limitations of bibliometric indicators as well as the time and effort for a bibliometric search for the total population of European HEIs, the EUMIDA Consortium decided to allocate extra resources (not included in the original budget) to a pilot study. It is a feasibility-in-the-feasibility study about the bibliometric profile of a sample of 57 HEIs drawn from the EUMIDA dataset.

Until recently, multidisciplinary bibliometric searches were exclusively performed by the database Web of Science (WoS) produced by Thomson Reuters. This database covers all fields of science and includes references (citations) in addition to pure bibliographic data.

This additional information enables scientific impact to be analysed. For some years now, the database Scopus – produced by Elsevier – is now available as viable alternative. Therefore the searches were conducted in both databases in order to obtain better insights into their comparative strengths and weaknesses.

3.4.1.2 Selection of HEIs for the bibliometric analysis

The HEIs for the analysis were selected from the 2,457 HEIs of the core data set. The general objective was to cover all countries and all types of HEIs. Furthermore, HEIs with a specific disciplinary focus should be included in order to investigate the publication intensity in different fields in an appropriate way. The selection was limited to HEIs which indicate that they are research-active – in total 1,405 HEIs equivalent to 47% of the core data set - as a relevant publication output could only be expected in these cases.
The practical investigation showed that for various HEIs – displayed as actively engaged in research - no or only very few publications could be identified, so that the researches had to be focussed on larger HEIs with a minimum total staff numbering about 1 000 persons. Furthermore, the criterion of granting a doctorate proved to be helpful. In the end, the sample focussed on public HEIs with a doctorate as the highest degree, which can be generally linked to HEIs primarily offering education at the ISCED 5a level. The other public HEIs and the private ones are not broadly covered. For Luxemburg, Malta and Cyprus only 1 HEI was covered; for most other countries 2 HEIs, for Germany, Spain, the United Kingdom and Italy 3 HEIs were analysed. In consequence, the sample is not representative, but it provides experience of the problems of obtaining institutional bibliometric data. In particular, the small number of institutions per country does not allow any reliable conclusions to be drawn as to the country coverage by the two databases. Therefore the institutional analysis was complemented by some countrywide searches.
Please note that this feasibility study has not any ambition in terms of representativeness of the sample. Rather, what is required is sufficient diversity to identify the main problems in bibliometric analysis.

3.4.1.3 Classification of publications by disciplinary area

The publication intensity – defined as the number of publications per researcher - largely varies by disciplinary fields (Schmoch and Schubert 2009). This observation is partly linked to specific publication practices by discipline, partly to the different coverage of disciplines in the publication databases. The two aspects cannot be clearly separated. Therefore it is not appropriate to exclusively look at the total publication numbers of HEIs, as this would favour HEIs with a focus on fields with high publication intensity. As more citations to other publications are possible in fields with a high number of publications, the fields with high publication intensity are also those with high citation intensity. In the study, the fields of

science were classified as indicated in Table 8. The classification by fields of science and technology (FOS) was introduced in the Frascati Manual in the 1960s and slightly revised since then. The last revision of FOS classification was conducted by the OECD in 2006 (OECD DSTI 2007).

Table 8. Fields of Science Classification

FOS 1	NATURAL SCIENCES	1.1 Mathematics 1.2 Computer and information sciences 1.3 Physical sciences 1.4 Chemical sciences 1.5 Earth and related environmental sciences 1.6 Biological sciences 1.7 Other natural sciences
FOS 2	ENGINEERING AND TECHNOLOGY	2.1 Civil engineering 2.2 Electrical engineering, electronic engineering, information engineering 2.3 Mechanical engineering 2.4 Chemical engineering 2.5 Materials engineering 2.6 Medical engineering 2.7 Environmental engineering 2.8 Environmental biotechnology 2.9 Industrial Biotechnology 2.10 Nano-technology 2.11 Other engineering and technologies
FOS 3	MEDICAL SCIENCES	3.1 Basic medicine 3.2 Clinical medicine 3.3 Health sciences 3.4 Health biotechnology 3.5 Other medical sciences
FOS 4	AGRICULTURAL SCIENCES	4.1 Agriculture, forestry, and fisheries 4.2 Animal and dairy science 4.3 Veterinary science 4.4 Agricultural biotechnology 4.5 Other agricultural sciences
FOS 5	SOCIAL SCIENCES	5.1 Psychology 5.2 Economics and business 5.3 Educational sciences 5.4 Sociology 5.5 Law 5.6 Political Science 5.7 Social and economic geography 5.8 Media and communications 5.9 Other social sciences
FOS 6	HUMANITIES	6.1 History and archaeology 6.2 Languages and literature 6.3 Philosophy, ethics and religion 6.4 Art (arts, history of arts, performing arts, music) 6.5 Other humanities

Source: OECD DSTI (2007)

For the bibliometric searches, a concordance of the classification of the documents in both databases and the 6 fields of science according to Table 8 had to be constructed. In most cases this is not problematic, but some specific issues have to be commented on.

In classifying of the *Fields of Education* for the EUMIDA survey, Social sciences and Education were separated. In the classification of the research-oriented *Fields of Science*, the Educational sciences are part of the Social sciences.

Computer and information sciences belong to the Natural sciences according to the FoS classification, in contrast, Information engineering to Engineering and technology. However, the classifications in publication databases are based on journals and not on single articles. In journals a strict separation between computer sciences and information engineering (software and hardware) cannot be realised. In these cases the journals have multiple classifications, so that fractional counting was applied.

Also in other fields, a clear distinction of Engineering and Natural sciences is difficult, for instance in Environmental or geological engineering. In this case, the publications were classified in Engineering.

A further borderline case is Radiology. This discipline is classified in the Medical sciences, although a relevant part of the publications deals with technical hardware. Nevertheless, also these articles are associated to Medical sciences, as a finer classification is not available.

Finally, the concordance scheme of Table 9 was constructed. In the case of WoS, each field is identified by a code of two characters, in the case of Scopus by a code of 4 digits.

Table 9. Concordance Scheme for the Fields of Science Classification for Bibliometric searches in WoS and Scopus

No	Field of Science	WoS		SCOPUS
1	Humanities	BP, BQ, EO, FS, HF, JS, JW, MM, MQ, OO, OR, OT, OX, OY, OZ, PA, PD, PF, PG, PH, QC, QD, QK, RP, UA, UT, YG, YI		1200, 3310
2	Social science	BM, BV, CN, DI, DK, EU, FE, FU, GY, HA, HB, HE, HI, JM, JO, MR, MW, MY, NM, NQ, NU, OE, OM, PC, PE, PS, UQ, UU, VI, VJ, VM, VP, VS, VX, WM, WQ, WU, WV, WY, XA, XW, YY, ZK		(3300 not (3310)), 1400, 1800, 2000, 3200 , 3304
3	Natural sciences	AA, BD, BF, BI, BU, CO, CQ, CU, DA, DB, DE, DR, DT, DW, DX, DY, EA, EC, EE, EI, EP, ER, ET, EV, EW, EX, FI, GC, GU, HT, HY, IY, JA, JB, KM, KU, KV, KY, LE, MC, MO, NS, OA, OU, PI, PN, PO, PQ, PT, PU, QQ, QU, RA, RE, RO, RQ, SA, SI, SY, TA, TE, UB, UF, UH, UI, UK, UN, UP, UR, WF, XE, XQ, XY, ZM, ZR		1000, (1300 not (1306,1308, 1310,1313), 1600, (1900 not 1909), 1700 (not 1708), (2300 not 2305), (2400 not (2403, 2404, 2405), 2600, 3100
4	Engineering and technology	IJ, IK, IL, IM, IO, IP, IQ, IU, IX, JY, PJ, PK, PM, PW, PZ, QE, QF, QG, QH, QJ, RB, RY, SR, UE, UY, YE, YQ, YR, ZQ		1500, 1708, 1909, 2305, 2100, 2200, 2500
5	Agricultural sciences	AD, AE, AF, AH, AM, JU, KA, MU		1100
6	Medical sciences	AQ, AY, AZ, BA, DM, DQ, DS, EQ, FF, FY, GA, GM, HL, IA, JI, KI, LI, LJ, LQ, MA, NE, NI, NN, OI, OP, PY, QA, RT, RU, RX, RZ, SD, SU, TC, TD, TI, TM, TQ, TU, UM, VE, VY, WC, WE, WH, YA, YO, YP, YU, ZA, ZC, ZD, ZE		2403, 2404, 2405, 2700, 2800, 2900, 3000, 3400, 3500, 3600, 1306, 1308, 1310, 1313

Source: EUMIDA

3.4.1.4 Coverage of fields by country

Analyses for articles in 2008 were conducted by country, based on the classification defined above. To illustrate the referring volume of publications, the results for WoS are documented in Table 10. It is obvious that the numbers for the fields of Humanities and Agriculture are very low, in particular, if the focus is on countries. In addition, the publication numbers in the Social sciences are quite low for smaller countries.

A further relevant aspect is the coverage by WoS compared to Scopus. To illustrate this issue, Table 10 records the relation of the number of publications in Scopus to WoS. Values of 1 indicate equal publication numbers, values above 1 a broader coverage by Scopus. On average, the relation is 1.2, thus the coverage in Scopus for European countries is about 20 percent higher than in WoS. Values above 1 are highlighted in bold letters. It has to be taken into account that in the presently available in-house version of Scopus (update June

2009) the country codes for many records are still missing. The provider announced a completion of these missing data for December 2010. Then a further increase of the numbers for European countries of about 6 percent can be expected.

Table 10. Number of Articles recorded in Web of Science for European Countries, 2008

Country	Humanities	Social sciences	Natural sciences	Engineering	Agriculture	Medicine	Total
Austria	133	489	4768	1090	142	3258	9880
Belgium	344	1081	6606	1692	237	4628	14588
Bulgaria	20	26	1514	285	86	295	2226
Cyprus	13	85	223	96	4	62	482
Czech Republic	98	318	4470	1042	245	1594	7767
Denmark	127	578	4401	873	218	3629	9825
Great Britain	2261	7444	30305	6700	485	22224	69420
Estonia	45	89	577	131	34	166	1042
Finland	124	654	4074	1053	265	2862	9033
France	1252	2277	31652	7354	821	14992	58347
Germany	1137	3817	40446	8451	968	23578	78396
Greece	69	430	3920	1605	184	3091	9300
Hungary	127	173	2973	502	410	1308	5492
Ireland	115	383	2368	679	123	1617	5284
Italy	525	1645	21891	5382	608	15085	45136
Latvia	1	30	231	100	11	59	432
Lithuania	106	175	793	481	87	299	1940
Luxembourg	2	36	141	46	2	107	335
Malta	12	8	48	13	0	41	122
Netherlands	436	2705	9779	2300	338	9926	25485
Norway	141	753	3415	764	263	2644	7980
Poland	133	329	9863	3543	477	3429	17774
Portugal	57	368	4207	1313	210	1206	7361
Romania	88	243	3173	1252	33	468	5258
Slovakia	88	136	1546	431	90	473	2763
Slovenia	125	210	1474	583	61	633	3086
Spain	807	2564	19487	5028	1080	9134	38099
Sweden	187	1124	7739	1787	258	6490	17586
Switzerland	214	1041	9212	1832	298	6180	18777
Total	7577	24480	194857	47762	6402	117674	398753

Source: EUMIDA

Table 11. Relation of the Number of Publications for European Countries in Scopus to WoS, 2008

Country	Humanities	Social sciences	Natural sciences	Engineering	Agriculture	Medicine	Total
Great Britain	0.66	**1.58**	**1.02**	**1.48**	**10.13**	**1.41**	**1.30**
Greece	**1.13**	**1.71**	0.93	**1.05**	**3.64**	**1.27**	**1.15**
Portugal	1.11	**1.50**	0.88	**1.12**	**4.00**	**1.45**	**1.14**
Ireland	0.41	**1.54**	0.90	0.92	**3.68**	**1.26**	**1.11**
Finland	0.68	**1.35**	0.93	**1.03**	**3.11**	**1.16**	**1.11**
Norway	0.61	**1.26**	0.88	**1.10**	**3.39**	**1.16**	**1.11**
Sweden	0.81	**1.35**	0.89	**1.20**	**4.88**	**1.15**	**1.10**
Spain	0.38	**1.07**	0.85	**1.07**	**3.74**	**1.39**	**1.10**
Netherlands	0.74	**1.17**	0.90	**1.17**	**4.60**	**1.14**	**1.09**
Czech Republic	0.49	0.91	0.83	**1.08**	**4.63**	**1.35**	**1.09**
Denmark	0.77	**1.31**	0.88	**1.05**	**4.24**	**1.13**	**1.09**
Bulgaria	0.95	**2.15**	0.90	**1.23**	**1.63**	**1.66**	**1.09**
Italy	0.61	**1.35**	0.91	**1.11**	**4.60**	**1.19**	**1.09**
Slovakia	0.83	**1.33**	0.86	**1.07**	**3.68**	**1.32**	**1.09**
Austria	0.57	**1.23**	0.90	**1.25**	**4.54**	**1.14**	**1.08**
Switzerland	0.53	**1.16**	0.93	**1.24**	**3.75**	**1.12**	**1.08**
Belgium	0.47	**1.14**	0.88	**1.09**	**4.82**	**1.17**	**1.07**
France	0.41	**1.29**	0.89	**1.25**	**4.40**	**1.14**	**1.05**
Hungary	0.70	**1.54**	0.87	**1.27**	**1.14**	**1.28**	**1.04**
Germany	0.54	**1.14**	0.91	**1.25**	**4.67**	**1.05**	**1.04**
Estonia	0.19	0.98	0.83	0.85	**3.92**	**1.48**	**1.03**
Luxembourg	0.73	0.90	0.87	0.87	**10.15**	**1.01**	0.98
Slovenia	0.22	**1.16**	0.88	**1.04**	**3.50**	0.98	0.98
Poland	0.58	**1.48**	0.82	0.78	**2.54**	**1.31**	0.96
Latvia	3.63	0.87	0.81	0.85	**2.36**	**1.15**	0.92
Romania	0.40	0.53	0.82	**1.10**	**3.71**	**1.15**	0.91
Lithuania	0.34	0.71	0.74	0.75	0.83	**1.09**	0.78
Cyprus	0.04	0.36	0.69	0.37	**3.61**	0.83	0.59
Malta	0.06	0.56	0.48	0.60	**6.81**	**0.72**	0.55
Total	0.59	**1.46**	0.98	1.23	**4.94**	**1.32**	**1.20**

Source: EUMIDA

However, there are substantial differences by field and country. A general observation is that the coverage of Humanities in WoS is much broader than in Scopus, but contrariwise the coverage of Agriculture is broader in Scopus. Among the disciplines with a high level of publications, the coverage of the Natural sciences in WoS and Scopus is nearly equal, with a slight advantage for WoS. In Engineering and Medicine, the coverage in Scopus is generally broader. These results refer to articles and the statements refer to the definitions of science fields according to Tables 8 and 9. There the attribution of specific sub-fields to either Natural sciences or Engineering is relevant. For instance, the classification of Information science is quantitatively relevant, as it is a large field with a better representation in Scopus.

The analyses by country cover all types of institutions, not only HEIs. Furthermore, they refer to articles. The integration of different document types in the analyses can make a significant difference. In Engineering, in particular, conference contributions are relevant types of publication. For instance, electrical engineering is covered more broadly in WoS, if only articles are considered. If conference contributions are also included, the coverage by Scopus is generally better.

For 21 of the 29 countries the coverage in Scopus is broader than in WoS, in further 3 cases, the level coverage in Scopus is at least at 95 percent of that in WoS (Table 11). Only in countries with a low absolute volume of publications is the coverage in WoS better than in Scopus, which can be due to specific preferences of some universities for journals not covered by Scopus.

The coverage of countries in the present version of Scopus is incomplete, and in many records, the country codes are missing, as the database is still being constructed. It can be assumed that in later versions, maybe already in 2011, the numbers for European countries will increase by about 6 percent. Thus the coverage by Scopus compared with WoS will be even better then. However, some activities of WoS can be observed which aim to enlarge the coverage of the database as well.

This general comparison at the country level suggests that bibliometric searches should be performed in Scopus, as the only major disadvantage compared with WoS is the low coverage of the Humanities, but in WoS the figures are also so low that they cannot be used for reliable statistical analyses.

3.4.1.5 Specialisation profiles of HEIs

Looking at the total number of publications by discipline it proves problematic to assess the country profiles in an appropriate way, as the countries differ considerably in size and in addition the publication intensity by field differs as well. A frequent solution for this issue is to calculate specialisation indexes RLA (Relative Literature Advantage). The corresponding equation is:

$$RLA_{ij} = 100 \tanh \ln [(Publ_{ij} / \sum_i Publ_{ij}) / (\sum_j Publ_{ij} / \sum_{ij} Publ_{ij})]$$

Here i stands for the country and j for the field. The RLA index is so constructed that its scale of values encompasses ± 100 with the neutral value 0. Positive values indicate an above-average specialisation, negative values a below-average one, whereby the European average serves as a reference in the present case. Values above +15 indicate a specialisation distinctly above average, while values below -15 are distinctly below average. A high value cannot be interpreted as high performance in absolute terms, but as a relative specialisation within the country's portfolio.

The results of such a calculation are presented in Table 12. In particular, the indexes in the Humanities and Social sciences have to be interpreted with care, as the samples are quite small and the adequate representation of the fields in the database is not verified. The indexes calculated on the basis of articles covered by WoS are similar.

Table 12. Specialisation Index RLA of European Countries based on Articles covered by in Scopus, 2008

Country	Humanities	Social sciences	Natural sciences	Engineering	Agriculture	Medicine
England	-27	-28	0	3	-9	7
Greece	11	5	-7	-5	10	7
Portugal	-17	-82	33	16	-14	-44
Ireland	-92	35	28	0	-28	-53
Finland	-47	-66	9	7	61	-24
Norway	-2	-5	-10	-35	26	17
Sweden	51	50	-16	-12	-19	7
Spain	-14	9	12	-16	55	-33
Netherlands	-10	16	-5	-13	22	2
Czech Republic	-10	-43	13	19	-12	-16
Denmark	-22	-33	12	5	-18	-7
Bulgaria	-24	-9	-17	24	-6	12
Italy	46	-45	12	-10	21	-11
Slovakia	-16	28	-10	-15	15	6
Austria	-35	-46	1	-1	-15	12
Switzerland	-2	-14	16	50	2	-56
Belgium	73	9	-3	58	-31	-38
France	-63	27	-7	0	9	1
Hungary	14	-12	-16	-9	-59	29
Germany	21	40	-24	-24	-17	22
Estonia	5	35	-17	-25	40	6
Luxembourg	-63	-75	16	27	7	-21
Slovenia	-21	-13	10	34	39	-42
Poland	-24	-77	29	69	-74	-78
Latvia	74	-22	10	22	47	-42
Romania	-2	7	7	45	7	-42
Lithuania	-25	-13	-1	4	36	-7
Cyprus	-17	4	-12	-11	-2	17
Malta	-47	-22	5	-9	-18	6

Source: EUMIDA

All in all, the profiles exhibit no clear structural difference between the large regions of West, East, North and South Europe. Rather, there is a distinct difference between countries with a strong orientation towards medicine and a weak one towards the natural sciences and engineering or vice versa, for instance, Germany and Hungary in contrast to Poland and Switzerland. Of course, in very small countries, the country profile can be largely equivalent to the profile of its largest university.

3.4.1.6 Limitations of this analysis

The basis for the Humanities and Social sciences proves to be limited in both databases, although the Social sciences are more broadly represented in Scopus than in WoS. In particular, many national journals, which are quite relevant in these scientific fields, are not included in the data sets. Furthermore, monographs and book contributions are also not covered, although they are quite important in these fields (Hicks, 2004).
In Engineering, the representation is better, but still needs to be extended. Here again, national journals are relevant, but also conference contributions (proceedings). In Mechanical engineering, the coverage of proceedings in Scopus is a little bit better than in WoS, here the major difference is due to a better coverage of articles in Scopus (Figure 5). In contrast, the coverage of proceedings in Electrical engineering in WoS is broader than in Scopus, implying an equivalent total number of publications in this sub-field (Figure 5). In Civil engineering only a very small number of proceedings could be identified in Scopus, although it covers a large number of proceedings in Information science. Thus the coverage of articles and proceedings is different by sub-field, so that overall statements for broader fields may hide these sub-structures.

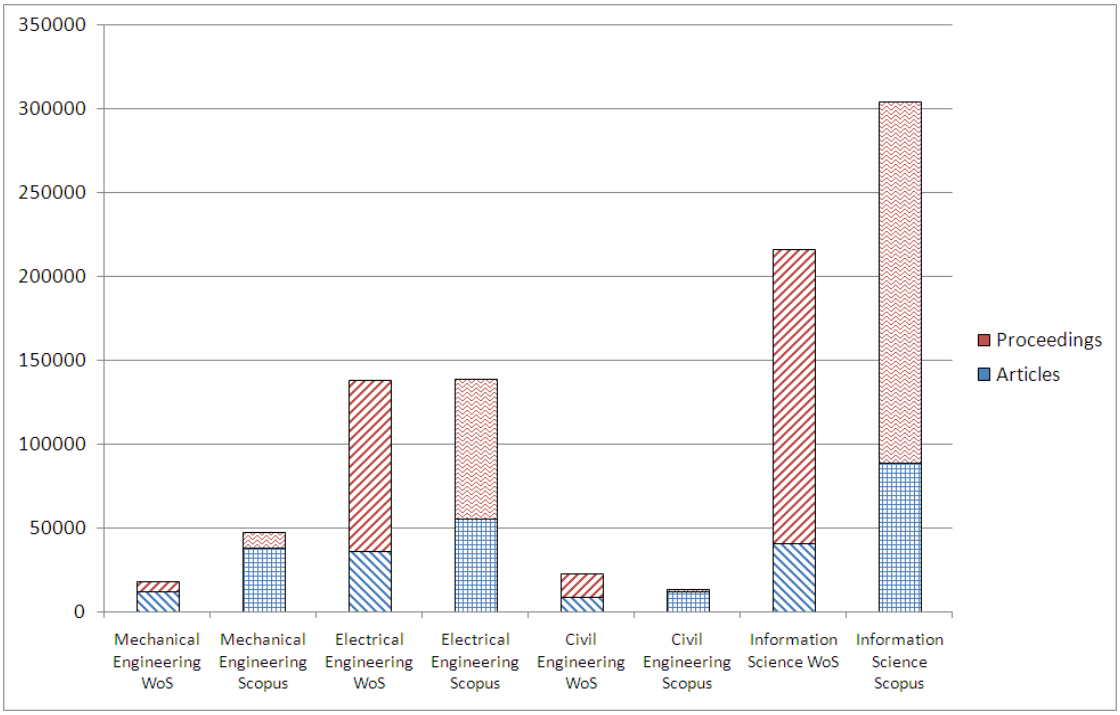


Figure 6. Number of Publications for selected fields, 2008 (Numbers for the total databases)

Source: EUMIDA

3.4.1.7 Findings at the level of individual HEIs

The HEIs for this analysis were selected according to the general criteria described above. All in all, 57 HEIs were included. This sample proved to be quite small and arbitrary within the whole population of research institutions, as shown in Table 13. The sample represents about 12 per cent of all publications in WoS for European countries and about 10 percent of all publications in Scopus for European countries. Thus for the HEIs of the sample, the relative representation in Scopus is lower than in WoS, i.e. many HEIs of the sample have a stronger orientation on journals exclusively covered by WoS.

Table 13. Sum of the publications of the selected HEIs in 2008 in the Scopus and WoS databases

	Humanities	Social sciences	Natural sciences	Engineering	Agriculture	Medicine	Total
WoS	657	3684	21365	4616	1008	15119	46448
Scopus	376	3693	18746	5060	3217	17514	48605
Relation Scopus/WoS	0.57	1.00	0.88	1.10	3.19	1.16	1.05

Source: EUMIDA

The sample is relatively small, as the analyses focussed on HEIs in medium-sized cities in order to facilitate the identification. In many large cities, several HEIs with only minor differences in their names are located, so that a strict separation can be problematic. But the researches proved sometimes to be complex, even for medium-sized cities. In a comparison of Scopus and WoS, the searches with Scopus were more effective and reliable, as the names of the institutions derived from the articles in the journals are not modified. This leads to a high variation of the institutional names, but the database producer introduced institutional identifiers (affiliation IDs). Thus the identity of the institution, as documented in the original name, is not modified, but the searches supported by the IDs were easier and more precise. In the case of WoS, the producer reworks the institutional names in order to reduce the variation of names and facilitate name-based institutional searches. However, the processors of WoS seem to be less familiar with European institutions and often "merge" separate institutions with similar names. As the original name version in the paper is not documented, it is impossible to correct these mistakes manually. In the present analysis, these problems were detected by a substantially unrealistic higher level of the WoS results in comparison to those in Scopus. The cases were eliminated and replaced by new ones.

Table 14. Publications and citations for the HEIs of the selected sample in Scopus and WoS, 2006

	Humanities	Social	Natural	Engineering	Agriculture	Medicine	Total
WoS publs	395	2194	20167	4200	566	13170	40692
Scopus publs	304	2672	15705	4446	2462	14554	40143
WoS cits	230	6564	120773	16618	2172	80075	226431
Scopus cits	207	7730	97653	18719	11726	114242	250276
Citation rate WoS	0.58	2.99	5.99	3.96	3.83	6.08	5.56
Citation rate Scopus	0.68	2.89	6.22	4.21	4.76	7.85	6.23

Source: EUMIDA

For the analysis of the citation rates, researches for the year 2006 were conducted with a 3-year citation window (2006 to 2008). As expected, the citation rates of WoS and Scopus are similar, but those of Scopus are a little bit higher, as the publications are cited by a higher number of publications due to the broader data coverage of Scopus (Table 14). But these differences in citation rates are so small that they are not a relevant criterion for choosing either WoS or Scopus.

In both databases, the citation rates in the Social sciences and the Humanities prove to be much lower than those in other fields. The citation rates in Medicine seem to be the highest ones. In any case, any assessment by the absolute level of citations or by non-standardised citation rates appears to be misleading. This problem can be illustrated by a HEI of the sample (Table 15). In this case, the publication numbers in Medicine are much higher than

in the other fields. With regard to the average shares of the publications in the different fields in Europe, a special orientation to Medicine is visible, but also a similarly strong orientation to the Social sciences, although the number of publications in the Social sciences is about one tenth of those in Medicine (specialisation indexes above +20). In terms of citation rates, the Natural sciences seem to be the strongest field in the HEI. But it is necessary to compare the observed rates to expected ones. We may define the citation rates for the total sample as reference. Then the relation of the rates of the HEI to the reference, the so-called Scientific Regard (Grupp *et al.* 2001), leads to a different picture. There the Humanities and Agriculture appear to be quite strong in terms of impact, Medicine has only an average level and the Natural sciences are clearly above average. To conclude, the definition of appropriate references proves to be quite important.

Table 15. Citations and Publications of a selected HEI in Scopus, 2006

	Humanities	Social sciences	Natural sciences	Engineering	Agriculture	Medicine	Total
Inst publications	21	191	511	102	25	909	1759
Specialisation	-12	36	-27	-49	-93	39	-
Inst citations	39	618	4781	750	246	7204	13639
Inst cit rate	1.83	3.24	9.36	7.33	10.01	7.92	7.75
Euro publications	304	2672	15705	4446	2462	14554	40143
Euro citations	207	7730	97653	18719	11726	114242	250276
Euro cit rate	0.68	2.89	6.22	4.21	4.76	7.85	6.23
Scientific Regard	**2.70**	**1.12**	**1.51**	**1.74**	**2.10**	**1.01**	**1.24**

Source: EUMIDA

The example for this relatively large HEI illustrates that the absolute numbers of publications in the Social sciences, Humanities and Agriculture are very low, so that the assessments based on this data is statistically not reliable. Looking at the total number of publications of the selected HEIs, the standard deviation in relation to the average number of publications for the years 2006 to 2008 was considered. For HEIs with an average number of publications below 200, the standard deviation was 20 per cent or even more. Above this value, the larger HEIs achieved values of 3 to 5 per cent, the medium-sized ones about 10 per cent. Broken down by fields, the absolute numbers for Humanities, Social sciences and Agriculture per institution are generally below 200, implying a permanent strong change of the citation rates over time. This low level of publications per field for single institutions is well illustrated in Table 16. Here the total number of publications per HEI covers a span between 33 and 4270 (26 and 3800 in Scopus).

Table 16. Publications in WoS of the selected HEIs sorted by the total Number, 2008[19]

No	Humanities	Social sciences	Natural sciences	Engineering	Agriculture	Medicine	Total
1	90	280	2253	271	74	1302	4270
2	72	256	1423	371	68	918	3107
3	28	134	1311	168	5	919	2565
4	27	134	1243	351	54	734	2543
5	41	273	917	123	6	1068	2427
6	33	131	1046	122	106	985	2423
7	26	94	688	62	23	1100	1993
8	13	154	606	61	11	995	1839
9	20	105	782	263	46	327	1543
10	8	251	214	29	1	968	1472
11	32	101	621	60	27	599	1439
12	6	97	577	267	9	379	1335
13	4	55	704	238	26	302	1330
14	13	35	463	36	1	724	1271
15	26	96	578	146	51	323	1220
16	12	52	532	115	12	402	1126
17	2	48	711	279	31	42	1113
18	17	451	246	8	191	180	1093
19	7	22	526	114	12	260	940
20	3	40	386	72	16	363	880
21	2	14	433	58	2	289	798
22	5	11	301	74	80	216	686
23	2	11	476	135	5	30	659
24	9	9	362	31	19	215	645
25	16	31	336	108	2	135	628
26	2	6	403	58	9	132	609
27	24	50	331	53	10	125	593
28	4	15	405	13	2	125	564
29	6	39	209	149	5	146	554
30	0	2	343	126	2	25	497
31	16	33	203	25	10	201	488
32	9	32	258	43	5	136	482
33	13	295	58	17	0	95	478
34	7	10	252	17	6	120	412
35	1	31	140	218	2	7	399
36	5	14	171	27	40	15	271
37	4	37	97	90	5	27	259
38	7	18	164	54	3	12	257
39	2	4	97	25	0	5	133
40	2	29	37	17	2	31	118
41	1	6	38	11	1	56	113
42	2	24	48	5	0	18	98
43	0	0	72	4	4	4	84
44	0	1	70	4	4	1	79
45	0	3	24	40	1	1	69
46	27	1	28	6	0	4	64
47	0	10	30	5	2	9	56
48	2	26	10	2	0	10	49
49	1	3	32	9	0	3	48
50	3	14	16	2	0	11	45
51	0	40	3	1	2	0	44
52	3	18	21	0	0	1	43
53	3	10	18	7	1	5	43
54	1	24	11	4	0	2	42
55	1	0	25	16	0	0	41
56	0	1	4	2	17	15	38
57	3	6	13	8	0	4	33

Source: EUMIDA

[19] The institutions with the number 5 is analysed in Table 15 as to the citations in Scopus referring to publications in 2006.

Table 17. Specialisation Profiles of the selected HEIs in Scopus with reference to the average distribution of all selected HEIs ordered by similarity

No	Humanities	Social sciences	Natural sciences	Engineering	Agriculture	Medicine
1	-90	-90	-6	-32	76	3
2	62	-26	-6	33	63	-33
3	-100	-100	-84	-52	87	49
4	49	-5	-12	-84	48	22
5	-45	12	30	-16	82	-96
6	-2	11	10	-45	27	-7
7	-100	-100	31	-82	88	-66
8	29	33	-1	-39	52	-12
9	-11	21	-5	-10	39	-7
10	83	-61	-18	85	-88	-68
11	-19	-11	-13	88	-58	-94
12	-47	-77	34	64	-19	-89
13	-40	-13	5	66	-37	-49
14	91	24	7	85	-100	-100
15	-49	-39	1	28	31	-14
16	-49	-9	8	27	20	-27
17	-27	4	26	44	3	-73
18	-100	-94	32	73	-89	-83
19	-100	-94	45	72	-100	-98
20	98	-92	24	41	40	-85
21	60	73	-37	64	14	-53
22	-80	-78	3	-29	-75	30
23	-45	-65	-26	-90	-68	54
24	-76	-12	-8	-36	3	22
25	-44	-77	12	-66	62	0
26	-18	-42	-15	-87	-25	43
27	-68	-55	-4	-15	1	20
28	-44	-23	-3	-56	-10	26
29	-22	-31	3	-40	-22	18
30	-40	-6	-11	-79	33	27
31	-41	-16	-12	-81	33	30
32	-72	-60	-74	-97	53	62
33	-41	0	-17	11	-41	21
34	30	-43	-15	-51	-9	35
35	11	-64	9	-28	-60	18
36	-37	23	-24	45	-81	11
37	-12	36	-27	-49	-93	39
38	-22	56	-84	-96	-89	65
39	-90	-80	50	9	-77	-78
40	-100	13	41	-24	-85	-53
41	-35	-66	33	-62	-11	-16
42	-100	-93	41	-50	54	-55
43	-52	-3	18	8	-59	-18
44	-46	-21	31	63	4	-95
45	-37	-18	1	6	-4	2
46	-27	-69	7	5	-30	9
47	73	92	20	-97	8	-99
48	-78	82	30	-63	-96	-68
49	-100	97	-58	-98	-78	-49
50	40	90	-59	-86	46	-2
51	71	88	-9	48	-90	-87
52	57	39	10	-48	-3	-12
53	96	87	-75	-86	59	3
54	62	85	-48	9	-22	-9
55	77	97	-85	-88	-98	-44
56	91	97	-70	-59	-24	-98
57	84	97	-47	-37	-100	-96

Source: EUMIDA

Most of the HEIs officially claim to cover all fields of science. This statement can be checked by a specialisation analysis by institution, as performed in Table 12 for the different countries. On this basis, most HEIs appear to be specialised in one or two fields, whereas the activities in the other ones are at a lower level (Table 17). Therefore it is possible to classify the HEIs by their major focus.

In order to obtain an impression of the publication intensity in the different fields, the number of the total publications per staff and HEI was calculated. Based on the results for specialised HEIs, the publication intensity in the different fields can be assessed. The numbers in Table 18 are very rough estimates, as the figures vary extremely between the different HEIs. This may be due to the selection of specific journals covered or not covered by the database, or a very strong mixture between different fields, or the different structure of the staff (the analysis is based on the total staff numbers, not only those for the academic staff). In any case, the low publication intensity in the Humanities and Social sciences compared to the other fields is obvious. The high score in Agriculture may be linked to the small size of this field and the related statistical problems, but the scores in Medicine are definitely the highest ones. This analysis shows again that publications must be analysed with reference to specific fields.

Table 18. Typical Score of the Number of Publications with Reference to the Staff Number of a HEI in the selected Sample

Field	Score
Humanities	0.04
Social sciences	0.06
Natural sciences	0.20
Engineering	0.20
Agriculture	0.60
Medicine	0.60

Source: EUMIDA

As the profiles of the selected HEIs are quite different, it was not possible to derive any reliable conclusions as to the differences by country.

3.4.1.8 Bibliometric searches for a larger population of HEIs

All in all, bibliometric analyses are useful to describe the research output of HEIs. However, relevant results can only be expected for larger, research-active institutions with a minimum level of staff of about 1,000. Due to these restrictions, bibliometric researches should be conducted for only about 1,000 HEIs in Europe. Only one database, either WoS or Scopus

should be used for these searches, if time and costs should be limited. However, the parallel search in both databases has various advantages as well.

The coverage of publications by Scopus generally appears to be broader than in WoS, at least in the majority of fields. A further argument for the choice of Scopus is the effort needed for the appropriate identification of a specific HEI. In the case of WoS, the institutional name is modified by the producer, with the aim of unifying names and facilitating institutional searches. For instance, the term "university" is changed to "univ". This "intellectual" modification works well for US institutions, as the producer of WoS, Thompson Reuters, is US-based. However, in the European context, many institutional settings are not appropriately understood, in particular in Eastern European countries with non-Anglo-Saxon language roots. In cities with several HEIs the sometimes slight difference in the institutional names are overlooked, HEIs located in cities with similar names are merged. This problem was detected in the present analysis by the obviously much too high publication numbers in the search results in WoS compared to Scopus.

In the case of Scopus, the identification of institutions is supported by institutional identifiers which are correct in the large majority of cases, but not always. In any case, they facilitate the searches.

There are various reasons to use both databases in parallel. First, WoS has a longer tradition than Scopus and a search including WoS will have a broader acceptance of the universities and the relevant institutions in the countries. Second, there are many HEIs with a broader coverage by WoS compared to Scopus despite the broader general coverage by Scopus. Thus a parallel search in WoS and Scopus would be the optimal strategy. The fuzziness of some search results as to institutions in WoS can easily be detected by a comparison of search results in both databases. The additional costs of a search in WoS are limited, as the majority of work is linked to the appropriate identification of HEIs by external information (Internet, HEI directories etc.)

The searches for specific institutions generally start with the identification of all institutions located in a specific city. Then the name variants of the target institution are taken up, in the case of Scopus the institutional identifier as well. Via a web search it must be verified whether the institution has affiliations at other locations and whether it is linked to a university hospital with a different institutional name. These searches can be very easy for small cities, but in larger cities, the correct identification may be complex. Particularly in very large cities with many HEIs such as Paris, Madrid, London, Rome, Vienna, or Berlin, correctly distinguishing the different institutions may be extremely time-consuming. On average, about 2 institutions can be identified per hour by an experienced researcher. But in the case of very large cities, about 3 days for clarifying these cases are realistic. Furthermore, about 1 month must be invested in planning the search process, implementing standard search strategies for time series and citations, and in compiling the search results. In addition, it will be necessary to arrange small consulting contracts with experts for the different countries in order to achieve an exact identification of institutions in more complex cases.

Finally it will be useful for the acceptance and the correctness of the results that the name variants of the institution and its affiliations, used for the searches, are verified by the HEIs themselves.

In total, the complete compilation of bibliometric data in Scopus will cost substantially less than € 100,000 for the first computation; a parallel search in Scopus and WoS will cost a little bit more than € 100,000. These costs include license fees for databases. In the following years, the costs will be about half the outlöays, as the major strategies for identifying institutions are already in place, and it must be only checked whether changes of the names or affiliations have happened.

Costs in this order of magnitude can only be realised, if the searches are limited to publications with clear institutional names. In some cases, only the names of the institutes, not the superordinate HEIs as such, are available. Sometimes only addresses are recorded. In these cases the related HEIs cannot be identified within a reasonable amount of time. Furthermore, crucial problems such as the association of French publications either to universities or to the CNRS cannot be solved.

The searches should be differentiated by fields. However, the data for Humanities, Social sciences and Agriculture should only be used to determine specialisation profiles, not for any institutional comparisons, as the numbers are too small for statistically reliable analyses and may extremely vary by year, and in the case of Humanities and Social sciences, other types of publications than journal articles are relevant and not covered by the available databases.

Despite these restrictions it is recommended to perform such an analysis as basic search for about 1,000 HEIs and to refine and improve the searches in the following years based on the experiences collected.[20]

3.4.2 Webometrics

As potential indicators for the analysis of the research output data from the Internet are suggested as an alternative or mainly complementary to bibliometric data. The methodological basis of this type of analyses – called webometrics – is well described by Björneborn and Iversen (2004). The most elaborated approach as to the assessment of universities is the *Webometrics Ranking of World Universities* performed by the Centro de Ciencias Humanas y Sociales (CCHS) of the Consejo Superior de Investigaciones Científicas (CSIC) in Spain. It covers about 12,000 universities worldwide, thus all important European universities are included. The authors describe their methodological approach as follows:

[20] Certain data included in this Chapter are derived from the Science Citation Index Expanded (SCIE), the Social Science Citation Index (SSCI), the Arts and Humanities Citation Index (AHCI), the Index to Scientific and Technical Proceedings (ISTP), and the Index to Social Sciences & Humanities Proceedings (ISSHP) (all update June 2010) prepared by Thomson Reuters (Scientific) Inc. (TR®), Philadelphia, Pennsylvania, USA, USA: © Copyright Thomson Reuters (Scientific) 2010. All rights reserved. Further data is derived from the in-house version of Scopus (status May 2009) prepared by Elsevier, Amsterdam, the Netherlands.

"The unit for analysis is the institutional domain, so only universities and research centres with an independent web domain are considered. If an institution has more than one main domain, two or more entries are used with the different addresses.

The first Web indicator, Web Impact Factor (WIF), was based on link analysis that combines the number of external inlinks and the number of pages of the website, a ratio of 1:1 between visibility and size. This ratio is used for the ranking, adding two new indicators to the size component: Number of documents, measured from the number of rich files in a web domain, and Number of publications being collected by Google Scholar database.

Four indicators were obtained from the quantitative results provided by the main search engines as follows:

Size (S). Number of pages recovered from four engines: Google, Yahoo, Live Search and Exalead.

Visibility (V). The total number of unique external links received (inlinks) by a site can only be confidently obtained from Yahoo Search.

Rich Files (R). After evaluation of their relevance to academic and publication activities and considering the volume of the different file formats, the following were selected: Adobe Acrobat (.pdf), Adobe PostScript (.ps), Microsoft Word (.doc) and Microsoft Powerpoint (.ppt). This data were extracted using Google, Yahoo Search, Live Search and Exalead.

Scholar (Sc). Google Scholar provides the number of papers and citations for each academic domain. These results from the Scholar database represent papers, reports and other academic items.

The four ranks were combined according to a formula where each one has a different weight, but maintaining the ratio 1:1:

WEBOMETRICS RANK		
VISIBILITY (external inlinks) 50%	SIZE (web pages)	20%
	RICH FILES	15%
	SCHOLAR	15%

The inclusion of the total number of pages is based on the recognition of a new global market for academic information, so the web is the adequate platform for the internationalization of the institutions. A strong and detailed web presence providing exact descriptions of the structure and activities of the university can attract new students and scholars worldwide.

The number of external inlinks received by a domain is a measure that represents visibility and impact of the published material, and although there is a great diversity of motivations for linking, a significant fraction works in a similar way as bibliographic citation.

The success of self-archiving and other repositories related initiatives can be roughly represented from rich file and Scholar data. The huge numbers involved with the pdf and doc formats means that not only administrative reports and bureaucratic forms are involved. PostScript and PowerPoint files are clearly related to academic activities (http://www.webometrics.info/about_rank.html, status September 27, 2010).

The data collected in the Webometerics project are impressive. In the present context, the question must be examined whether this data could or should be used as alternative or complementary indicators for research output. For this purpose some reflections on methodology are useful. As a major point the problem of the adequacy of indicators (Grupp 1998) has to be examined. Adequacy refers to the problem that indicators are always a proxy for a phenomenon that cannot be measured directly. Then it is important that the indicator represents the phenomenon – here the research output – in an adequate way, thus that it what is measured by the indicator is clearly defined.

In the case of Webometrics, a first methodological problem is that the key indicator is a composite indicator combining four different sub-indicators with different weights. Therefore different aspects of output are mixed and it is not clear what the meaning of the rank defined by the combined indicator really means. This issue may be illustrated by the ranking of the top 10 European universities documented in Table 19. For instance the first European university, the University of Cambridge, with the world rank 22 has good positions first of all in the dimensions Size and Visibility whereas the ETH in Zürich which ranks second in Europe is much better in the dimension Scholar or the University of Oslo in the composite rank 5 in Europe is better in the dimensions Rich files and Scholar. So the reasons behind a certain rank are fuzzy and obviously depend on the weight of the different dimensions.

Table 19. Extract of the list of top universities 2010 according to the CCHS-CSIC project

CONTINENT RANK	UNIVERSITY	COUNTRY	WORLD RANK	SIZE	VISIBILITY	RICH FILES	SCHOLAR
1	University of Cambridge		22	15	18	54	97
2	Swiss Federal Institute of Technology ETH Zürich		40	38	57	73	15
3	University of Oxford		41	55	30	68	115
4	University of Edinburgh		52	96	50	63	86
5	University of Oslo		53	58	98	35	46
6	University of Helsinki		59	73	87	48	55
7	Norwegian University of Science & Technology		61	40	117	71	26
8	University College London		63	107	69	70	72
9	Universität Wien		69	82	118	92	18
10	École Polytechnique Fédérale de Lausanne		72	30	128	83	64

Source: http://www.webometrics.info/top100_continent.asp?cont=europe

A second problem is that for the sub-indicators Size, Visibility and Rich files a relation to education as well as research is possible. Many HEIs publish documents for teaching

purposes in the net for their students – assuming that all students have a computer with internet access - so that these three indicators reflect also education to a certain – unclear – extent. In any case they are not a distinct output of research. In the case of the dimension Scholar, the vague basis of Google Scholar – mix of different types of documents, mix of different disciplines, a steadily changing basis of reference – is critical. Google Scholar may be useful, if it could be used in a differentiated way by distinguishing disciplines and types of documents, but for that purpose additional work has to be invested to achieve improved differentiated indicators. In the present situation analyses with bibliometric databases are more precise and adequate.

An interesting sub-indicator is Visibility, as it reflects the aspect of transfer in a new perspective, by collecting the number of inlinks of a website of a HEI. For the systematic use as an indicator for research output, further studies are necessary on which institutions are referring to a website and which types of content they refer to. Then it would be possible to distinguish different types of inlinks or at least to determine the weight of these different types.

All in all, the indicators provided by the *Webometrics Ranking of World Universities* are interesting for integration in a broader collection of research output indicators for European HEIs, but further work is necessary to achieve an appropriate relation between some sub-indicators, in particular Visibility, and research output.

3.5 Conclusions

With regard to data on research outputs, the present collection of the EUMIDA data set provides useful information on students at the ISCED 6 level already available in the core data set and differentiated by fields of education in the extended data set. This information proves to be particularly relevant for the fields of Social science and Humanities as well as for smaller HEIs, which are not sufficiently covered by bibliometric indicators. From the data on ISCED 6 students it is also possible to derive information on internationalisation.

The extended EUMIDA data collection provides data on R&D expenditure for about 40% of the research-active HEIS that represent about half of the ISCED 6 students of these institutions. An interesting piece of additional information as to the transfer orientation of HEIs is the share of private funding of R&D. Some validity checks show that there are some inconsistencies in this data, but at quite a low level. The major limitation of this data is that the HEIs of some countries, where a specific investigation of HEI research is not performed and the public funding for education and research is not separated, are systematically excluded. As this data is very useful for policy in higher education, all countries should be encouraged to change their inquiry practice.

The availability of data on academic patent and spin-off companies based on surveys proves to be quite limited and not useful for broader systematic comparisons. It is recommended to use approaches where names of academic staff and inventors are matched. In particular,

the use of bibliometric databases for generating academic name lists proves to be promising. But the related research is not completed yet and results will only be available in Spring 2011. Alternative approaches are not suggested for spin-off companies.

A feasibility study revealed that bibliometric searches for larger research-active HEIs are possible and can be realised for all these HEIs for reasonable outlays in time and costs. In particular, they are useful for the fields of the Natural sciences and Medicine, but also – with certain limitations - for Engineering. Their relevance for the Social sciences and Humanities is restricted due to a narrow coverage of journal articles and a complete exclusion of other relevant types of publication (monographs, book contributions, national articles). The optimal approach is a parallel search in the databases WoS and Scopus.

At the present status of research, a use of webometrics for the generation of alternative data on HEIs is not recommended, bur rather its use as a complementary source. The dimension of Visibility appears to be promising, but more analyses are necessary to clarify which aspects of transfer these indicators reflect.

All in all, the presently available data on research output for specific HEIs are limited and should (and could) be complemented by patent and publication data. Webometric data may be useful in the long run, but not at the present stage of research.

4. Data availability and procedures for data collection

4.1 Introduction

This document provides the methodological foundations of the EUMIDA data collection for the core and full set of data. It includes the procedures for data collection and sources presented in D3 and D4 - offering guidelines for both, core and extended data collection on European higher education institutions.

This document covers the following main topics:

- In Section 4.2, *national propositions for the perimeter* and some open issues are presented. A distinction is made between research-active institutions and others.
- In Section 4.3 *data availability, gaps and sources* for the core set of data are examined. Data availability and confidentiality in each country are described.
- In Section 4.4 *data availability, gaps and sources* for the extended set of data are examined. Data availability and confidentiality in each country are described.
- In Section 4.5 the main *actors and roles* involved in data collection are presented in detail, namely the data owners that could provide information in both the core and extended set for each country.
- Finally, in Section 4.6 the EUMIDA proposition for the *procedures for data collection* both in the EUMIDA project and for the operational phase are introduced.

The content of this document has been extensively discussed within the EUMIDA core team, with the European Commission, Eurostat and with the EUMIDA National Experts. Definitions, delimitations and special cases have been revised based on the feedback of National Experts, as well as on problems emerging from the pilot data collection in some countries.

4.2 National propositions for the perimeter

This Section presents a summary overview of national situations concerning the definition of perimeter in the EUMIDA countries; it is based on EUMIDA national experts reports and on deliverable D2 on data availability. In most countries, the definition of the perimeter, as well as the distinction between research-active institutions and other institutions, was rather unproblematic. The following Table 20 shows the number of institutions covered by both the core and extended set of data.

Table 20. Number of institutions covered in the EUMIDA project

Country	No. Institutions (Data Collection 1)	No. Institutions (Data Collection 2)
Austria	68	68
Belgium	87	43
Bulgaria	59	39
Cyprus	37	7
Czech Republic	73	26
Denmark (^)	n.a.	14
Estonia	34	7
Finland	49	49
France (^^)	n.a.	n.a.
Germany	410	306
Greece	60	40
Hungary	72	61
Italy	243	81
Ireland	21	21
Latvia	61	20
Lithuania	46	16
Luxembourg	1	1
Malta	4	1
Netherlands	59	59
Norway	68	44
Poland	457	91
Portugal	138	18
Romania	82	57
Slovakia	33	33
Slovenia	13	4
Spain	47	47
Sweden	49	41
Switzerland	36	36
United Kingdom	150	148

(^) Data from Denmark reflects the fact that this country has not delivered any data for DC1, but was able to deliver for 14 institutions in DC2.
(^^) Data from France are missing. According to the Country report, the number of institutions in the DC 1 would have been 435, of which 120 research active.

Thus the total number of institutions covered by the EUMIDA data collection is 2,457 for DC 1 (which excludes Denmark and France) and 1,378 for DC 2 (which excludes only France and includes 14 Danish institutions in DC 2).

In addition, if:

- France were included, following the estimate of the perimeter done by experts and indicated in the footnote (see also Table 21 below)
- The same number of institutions for Denmark identified in DC 2 were included in DC 1 (which is a conservative estimate)

then the overall European perimeter would consist of 2,906 institutions, of which 1,498 research active.

The following Table 21 summarizes currently available information per each country.

Table 21. Perimeter for core and extended set of data

Country	Institutions in DC1	N.	Institutions in DC 2	N.	Not in EUMIDA perimeter	N.	Remarks
Austria	Public Universities University of further education, Krems Private universities Universities of Applied Sciences (UAS) University college of teacher education	21 1 12 20 14	Public Universities University of further education, Krems Private universities Universities of Applied Sciences (UAS) University college of teacher education	21 1 12 20 14	School of theology Colleges for medico-technical occupations and colleges for midwifery Courses for technological, trade and commercial occupations; training Kollegs for kindergarten teachers; training Kollegs for non-teaching supervisory staff Courses for master craftsmen, foreperson and construction trades	3 21 around 60 65	Institutions for extended set of Data is the research active institutions
Belgium	Universities Hogescholen, Hautes Ecoles, Hochschule, Ecoles supèrieures des Ats, Instituts supèriurs d'Arquitecture.	14 52 21	Universities Hogescholen, Hautes Ecoles, Hochschule,	7 36	--	--	Institutions for extended set of Data are the research active institutions
Bulgaria	Universities Academies of sciences Independent colleges Research institutes Specialized Higher school	26 2 11 4 16	Universities Academies of sciences Research institutes Specialized Higher school	21 2 4 12	Colleges	38	Independent colleges are excluded in Data Collection 2 because they do not have a research mission.
Cyprus	Universities Institution of tertiary education	6 31	Universities Institutions of tertiary education	6 1	schools not recognised as tertiary education institutions by Ministry of Education	4	Institutions for extended set of Data is the research active institutions
Czech Republic	Universities Non universities institutions	28 45	Universities	26	Tertiary Professional schools	123	Only public universities are legally considered as institutions with a research mission and with relevant number of PhD students.
Denmark	--		Universities Business school Academies of art, school of design or architecture Royal school of library	7 1 5 1	--	--	data for DC1 has not delivered but where able to deliver for 14 institutions in DC2
Estonia	Universities Prof higher education institution Vocational education school	10 21 3	universities	7	--	--	Institutions for extended set of Data is the research active institutions

Country	Institutions in DC1	N.	Institutions in DC 2	N.	Not in EUMIDA perimeter	N.	Remarks
Finland	Universities	21	Universities	21	--	--	Institutions for extended set of Data is the research active institutions
	Polytechnics	28	Polytechnic	28			
France	Universities					81	The list of categories provided in the Table is extracted from the Country report from official sources. No official validation of the perimeter was done at NSA level. The guess of experts is that only 120 institutions might fall in the perimeter of DC 2 (i.e. the first six categories) and 435 in total in the larger perimeter of DC 1. However, this must be considered an expert opinion, not a validated set of figures.
	Grands Etablissements					10	
	Technical universities					3	
	Ecoles Normales Superieurs					4	
	National Polytechnics					2	
	Engineering Schools under MESR					20	
	Engineering Schools under other Ministry					39	
	Private engineering Schools					68	
	Business Schools					208	
	High schools proposing CPGE, excl. those in engineering and business schools					422	
	High schools proposing Higher technical education sections (STS)					2133	
	Other private HEI					13	
	Higher professional schools, incl. Arts and culture related schools					452	
	Higher professional schools					598	
Germany	Art colleges	52	Art colleges	45	--	--	Institutions for extended set of Data is the research active institutions
	Theological colleges	15	Theological colleges	9			
	Universities	111	Universities	109			
	Univ of applied Sci	192	Univ of applied Sci	133			
	Univ of Education	6	Univ of Education	6			
	Univ of Publ adm	34	Univ of Publ adm	4			
Greece	Universities	24	Universities	24	--	--	All HEIs (Universities and Tech. Educational Institutes) are included because these are legally considered as institutions with a research mission, and detailed institutional level Data is available
	Technological educational institutes and school of pedagogy	16	Technological educational institutes and school of pedagogy	16			
	Third level professional educational school	20					
Hungary	Generalistic universities	16	Generalistic universities	16	Universities & colleges and other institutions	30 (HEIs) + 143	Institutions for extended set of Data is the research active institutions
	Theological universities	3	Theological universities	2			
	Medical univ	1	Colleges	37			
	Colleges	47	Medical univ	1			
	Univ of art	4	Univ of art	4			
	Univ of defence	1	Univ of defence	1			
Italy	Universities	72	Universities	69	--	--	The perimeter includes HEIs legally considered as institutions with both a research mission consistent and stabilized research activities. Data is easily available. On the contrary AFAM Institutes and very
	Universities for foreigners	3	Higher school	6			
	Technical univ	3	Online university	1			
			Technical univ	3			

Country	Institutions in DC1	N.	Institutions in DC 2	N.	Not in EUMIDA perimeter	N.	Remarks
	Online univ Acad of fine art Free academy Higher inst for music and art Higher school	11 20 24 81 29	Univ for foreigner	2			young Universities are still at a preliminary level with the development of the research work.
Ireland	Universities Institutes of Technology	7 14	Universities Institutes of Technology	7 14	Colleges of Education Other Colleges	7 58	Institutions for extended set of Data is the research active institutions. Only institutions that award their own degrees are included
Latvia	Universities State HEI State colleges Other Inst in Higher education	6 13 18 24	Universities State HEI Other Inst in Higher education	6 9 5	--	--	It is proposed to include those 20 HEIs defined as research active based on the presence of a research mandate, inclusion in the R&D statistics and awarding doctorates.
Lithuania	universities colleges	23 23	universities	16	--	--	It is proposed to excluded the private HEIs due to the lack of microdata available from publicly accessible sources
Luxembourg	University of Luxemburg	1	University of Luxemburg	1	--	--	Stricto sensu, only one institution has the label of university in Luxembourg. Therefore, it is suggested to limit the perimeter of the data collection to this institution.
Malta	University of Malta Private institute Vocational college	1 1 2	University of malta	1	--	--	The University of Malta is the only research-active institution in Malta. Therefore, the extended set of data collection will only include the University of Malta.
Netherlands	Public universities Privates universities Universities of Applied Sciences (Hogescholen)	14 5 40	Public universities Privates universities Universities of Applied Sciences (Hogescholen)	14 5 40	Private institutions of professional education	58 (estimate)	Institutions for extended set of Data is the research active institutions
Norway	Universities Specialized university State university colleges Private univ college Police and military univ college	7 8 26 21 6	Universities Specialized university State university colleges Private univ college	7 8 26 3	Professional schools	Approx 100 tertiary education institutions (public and private)	The institutions included are legally considered as institutions with a research mission, research is a requirement for their accreditation and detailed institutional-level Data is available in the various relevant databases
Poland	Universities Technical universities Academies Other HEI Teacher educ school	18 24 151 246 18	Universities Technical universities Academies Other HEI Teacher educ school	17 18 48 3 5	--	-	In the extended set of data will be included institutions with more than 50 PhD students

Country	Institutions in DC1	N.	Institutions in DC 2	N.	Not in EUMIDA perimeter	N.	Remarks
Portugal	Public universities Private universities Public polytechnic Private polytechnic Public Military HEI	15 42 20 56 5	Public university Private universities	14 4	--	--	Institutions for extended set of Data is the research active institutions
Romania	Universities Technical universities	76 6	Universities Technical universities	51 6	accredited higher education institutions authorized to function provisory	24	It is proposed to include in the perimeter for extended data collection 41 state civil universities. There are 24 accredited HEIs authorized to function provisory which are not included in EUMIDA perimeter
Slovakia	Public Higher Education Institutions State Higher Education Institutions Private Higher Education Institutions	20 3 10	Public Higher Education Institutions State Higher Education Institutions Private Higher Education Institutions	20 3 10	--	--	Institutions for extended set of Data is the research active institutions
Slovenia	Publics and privates universities Single higher education institutions (colleges + faculties)	4 9	Publics and privates universities	4	Vocational colleges	59	HEIs with relevant educational and R&D activities will be included.
Spain	Public universities	47	Public universities	47	Private universities Private universities for distance education Public university for distance education Public special universities Professional High School	23 4 1 2 2071	Only public universities are included in DC due to the relevant research activity and availability of data
Sweden	Universities/HEIs University colleges of Arts Other university college	16 9 24	Universities/HEIs University colleges of Arts Other university college	16 7 18	--	--	Institutions for extended set of Data is the research active institutions
Switzerland	Cantonal universities and FIT Universities of Applied Sciences Teacher training institutions Other HEI	12 12 10 2	Cantonal universities and FIT Universities of Applied Sciences Teacher training institutions Other HEI	12 12 10 2	Federal professional exams High professional schools	500 institutions offering preparatory courses	Teacher training institutions are most small and with limited research intensity. Kalaidos Fachhoschule has been recently created, research-intensity is low and data availability might be problematic.

Country	Institutions in DC1	N.	Institutions in DC 2	N.	Not in EUMIDA perimeter	N.	Remarks
United Kingdom	Universities University college	131 19	Universities University college	131 17	Listed bodies (no power to award degrees; can deliver higher education degrees accredited by universities)	606	Institutions for extended set of Data is the research active institutions

4.3 Data availability, gaps and sources. Core set of data

According to national reports and D2, availability of data for the core set of variables is relatively unproblematic in most countries.
More in detail, the situation by indicator is as follows:

4.3.1 Availability

Identifier

Institutional code

No problem with this indicator in any country.

Legal name of the institution

Legal name and official English translation could be obtained in all countries. In Belgium, no official English translation exists for the name of HEI. Existing unofficial English translation can be obtained from the website of some of the institutions.

Institutional descriptors

Country

In all cases data for national campus is complete and available. There is incomplete information for foreign campuses in Czech Republic, Greece, Ireland and Malta.

Legal status

Considering public or private control, only in Austria is not readily available but it can be provided.

Foundation year and current status year

Foundation year is a complex indicator because the definition is not homogeneous in the different countries. In Bulgaria, Estonia, Ireland and Malta this information is not directly available and in Hungary there are only data for public university. It has to be checked to which extent having two dates will help solving this problem. Some difficult cases might remain. Otherwise, most data on *foundation year* and *current status* are easily available from the web site of each institution.

University hospital

Data about existence of *university hospital* is available in all countries.

Staff

Data on total staff are available in the majority of countries. In Portugal, Data is available only for public institutions. In Belgium, data for the non university HEI are available but not public (available and published for the universities).

Educational activities

Number of students

The number of *enrolled students* is available, in general, in all countries but in Netherlands only for public universities.

Specialization

Data about *specialization* are complex too because in different countries there are diverse classification and level of aggregation and different criteria to consider *specialization*. This data has to be more clearly defined. There are some problems with availability in Czech Republic, Denmark, Estonia, Ireland and Latvia.

Distance education institutions

Availability problems are not detected.

Highest degree delivered

Indicators *highest degree* is mainly available for all countries.

Research activities

Research active institutions

Research activity is an indicator available in the majority of country (not for Slovenia). In Romania NIS collects data for public and private universities but these are publicly available only in aggregated format Potential problems are detected in France with private and business schools.

Doctorates awarded

Doctorates awarded is only not available in Bulgaria and for private universities in Malta.

Knowledge exchange

No indicators yet

International attractiveness

Absolute number of international under graduated and *absolute number of international doctoral students* are available in the majority of countries.

Regional engagement

Region of establishment

Data about *region of establishment* are only not available in Ireland because the level of aggregation is different than in other countries. Such attribution needs to be made ad-hoc, as universities are not classified by HEA on a regional basis.

4.3.2 Reasons for non-availability (confidentiality)

For the core set of variables, reasons for non-availability seem to be relatively minor. Only problems with *number of enrolled students* in Bulgaria, Romania and Slovenia are detected and for *doctorates awarded* in Slovenia and Romania. In these two countries NIS collects the data regarding the number of enrolled student and doctorate for each public and privates universities but these are publicly available only in aggregated format. These could be available by unit only after individual agreement of the providers. For the other indicators

problems are not detected. In some cases there are less data available for private institutions, but more data could be obtained contacting directly with the owners.

4.3.3 Summary overview

Table 22 shows a summary overview of the state of data by country. Rather few problems concerning availability and reasons for non-availability are identified and thus feasibility of the core set data collection can be confirmed. In most cases, it appears relatively easy to collect missing data, especially in case of descriptors.

Table 22. Summary overview of availability and confidentiality by country

	Availability	Confidentiality
	Some difficulties in	Some difficulties in
Austria	Foundation year (in 5 institutions) ISCED 5 students (in 1 institution) number of total staff (in 43 institutions)	
Belgium	Foundation year (1 institution) Doctoral degree (3 institutions) ISCED 5 (45 institutions) International ISCED 6 (4 institutions) Total staff (51 institutions)	
Bulgaria	Foundation year (2 institutions)	number of doctorates (57 institutions) ISCED 5 (in 2 institutions) ISCED 6 (in 6 institutions) total staff 22 (institutions)
Cyprus		ISCED 5 (2 institutions) Total staff (2 institutions)
Czech Republic	Foreign campus, Specialization	
Denmark	Specialization	
Estonia	Total staff (34 institutions)	
Finland	ISCED 5 (1 institution) Total staff (3 institutions)	
France		
Germany	Legal status (2 institutions) Foundation year (1 institution) Highest degree (36 institutions) Doctoral degree (36 institutions) ISCED 5 (11 institutions) International ISCED 6 (30 inst) Total staff (7 institutions)	
Greece	ISCED 5 (1 institution) Total staff (1 institution)	
Hungary	Current status year (3 institutions) Distance education (2 institutions) Research active (1 inst) Univ hospital (1 institution) Highest degree (1 institution) Doctoral degree (4 institutions) ISCED 5 (2 institutions)	

	Availability	Confidentiality
Ireland	Foreign Campus, Foundation year, Region, Specialization	
Italy	Foundation year (49 institutions) ISCED 5 (27 institutions) Total staff (26 institutions9	
Latvia	Current status year (2 institutions) Total staff (1 institution)	
Lithuania		ISCED 5 (8 institutions) ISCED 6 (8 institutions) Total staff (8institutions)
Luxembourg		
Malta	Foundation year (2 institutions) Current status year (2 institutions) International Students (3 inst)	
Netherlands	International Students (5 inst)	
Norway	ISCED 5 (5 institutions) Total staff (6 institutions)	
Poland	Distance education (457 inst) ISCED 5 (2 institutions)	
Portugal	Current status year (1 institution) ISCED 5 (1 institution)	
Romania		ISCED 5 (7 institutions) ISCED 6 (6 institutions) Total staff (7 institutions)
Slovakia		
Slovenia		NUT region (2 institutions) ISCED 5 (2 institutions) ISCED 6 (2 institutions) Total staff (2 institutions)
Spain		
Sweden	International Stud (6 institutions)	
Switzerland	Foundation year (19 institutions) Total staff (1 institution)	
United Kingdom	ISCED 5 (3 institutions)	

Source: EUMIDA

4.4 Data availability, gaps and sources. Extended set of data

According to national reports and D2, availability of data for the extended set of variable is more complex than that for core set.

4.4.1 Availability

In general, the main problem of availability is that there are no official sources to obtain output information on Higher Education Institutions. The following provides more in-depth information on availability of data for each variable, as well as on the specific definition and methodological issues introduced in the EUMIDA handbook. The following tables present data on each category by country.

Expenditure

Data on expenditure is available for (virtually) all institutions in Germany, Italy, Netherlands, Spain, Sweden and UK (Table 23).

Coverage of expenditure data is different in each country but in all cases the coverage is better for public HEIs (Austria, Bulgaria, Czech Republic, Finland, Lithuania, Portugal, Slovakia and Spain).

Disaggregation of data between current expenditure (personnel-non personnel) and capital expenditure is not easy in the majority of countries. Expenditure in personnel data is in general available (except for Denmark) but only at aggregate level in Cyprus, Greece, Hungary and Romania). No systematic accounting of capital expenditure is detected in Estonia and it is not available in Switzerland since for many institutions it is still included in public budgets. There are potential difficulties for the separation between education and healthcare in university hospitals in some countries (such as Czech Republic, Estonia, France, Germany, Portugal and Switzerland) (see Table 24).

Table 23. Availability of data on expenditure by country (N institutions without data)

	Missing cases	Not available	Not applicable	Confidential	Unreliable or uncertain data	Deviates from definition	Provisional value	Total no. of institutions
AT				46				68
BE		13	1					43
BG	1	4		6				39
CH		5						36
CY		1						7
CZ		2						26
DE		5						306
EE				7				7
ES								47
FI		3						49
GR		40						40
HU	1	32	2					61
IE								21
IT		2						81
LT				16				16
LU								1
LV				5				20
MT								1
NL		1						59
NO								44
PL		92						91
PT		4						18
RO		4		1				57
SE								41
SI		1		2				4
SK		13						33
UK								148
Total	**2**	**222**	**2**	**83**				**1364**

Source: EUMIDA

Table 24. Summary overview of availability of data on Expenditure by country

Country	Coverage	Current expenditure: personnel	Current expenditure: non personnel	Capital expenditure	Breakdown: other	Remarks
Austria	Public universities	available but confidential, public provision currently discussed	available but confidential, public provision currently discussed	available but confidential, public provision currently discussed	depreciation	
Belgium						This information still has to be checked.
Bulgaria	Public universities; BAS ACA	Type of expenses: Salaries, Social security, Upkeep (food, running costs, business trips, etc.), Scholarships, Repairs, Assets	Available by cost categories	Available	Type of activities: Dormitory campus, Advance studies, Canteen, High schools, Applied research, International contracts, Other activities	NSI has data about all 59 institutions but considers it confidential. MEYS is publishing annual accounts for the past 3 years (it is already changed to 2007,2008,2009). i.e. the 2009 is available since a couple of weeks.
Cyprus	Private and public university and non-university level institutions	Available by Statistical Service at aggregate level for private institutions and disaggregated for public institutions	Available by Statistical Service at aggregate level for private institutions and disaggregated for public institutions	Available by Statistical Service at aggregate level for private institutions and disaggregated for public institutions	There is no other breakdown available.	
Czech Republic	Public HEIs of universities type	Available	By cost categories (supplies, energy, maintenance, travel cost, capital consumption, other)	Available	By activity (education and R&D, other activities).	Accounting of university hospitals is separated from universities´ budgets, but the funding of educational and research activities are in majority included in IIE
Denmark	Core university institutions: 8 universities and 6 university-like institutions	Not available	Not available	Not available	Education, research, general admin, facilities	The 8 universities produce a joint set of statistics. The indicated availability is based on this set. The remaining "university-like" institutions do not produce this kind of common figures, but there are chances that this may happen. Otherwise, data must be collected by contacting each institution.

Country	Coverage	Current expenditure: personnel	Current expenditure: non personnel	Capital expenditure	Breakdown: other	Remarks
Estonia	Public and private universities	Personnel expenditure (separately on teaching staff)	Available	No systematic accounting of capital expenditure.	The costs include current expenditure, investments and costs of auxiliary services for students. The costs of other services besides tertiary education are excluded. Costs on R&D are excluded.	The cost of medicine is contested because of difficult separation between education and healthcare in university hospitals (in the case of University of Tartu).
Finland	Universities and Polytechnics in the field of Ministry of Education and FI031	Universities: Budgetary funding (core funding) on salaries excluding third party. Polytecnics: core and third party, by type of education, research, regional development, premises maintanance, shared activities.	Universities: Budgetary funding on 1) premises 2) other operational expenditure. No data on third party. Polytecnics: Core and third party, by type of education, research, regional development, premises maintanance, shared activities. Including/excluding VAT	Universities: budgetary funding on building investments. Polytechnics: Capital expenditure included in internal rents reported as an item of premises maintanance expenditure. Investments reported in separate report, but not yet checked for double counting.	Universities: External financing by source, expenditure by performance area and education fields (not fully compatible with ISCED97)	Universities: this breakdown is not supported for externel funding. Total expenditure available using education/research/other - breakdown. Polytechnics: Board of education report K04K7SL. Expenditure do not cover services for fee, i.e. figures cover expenditure supported by state grants. Investments from report K94J6OL
France	All institutions of the extended data set	Available	Available	Available	By educational field (based on a ad hoc a classification of institutional units into fields). By LOLF actions breakdown	Changes in the technical aspect of reporting (towards more automatization and more reliability). LRU 2007 law entails changes in reporting system and perimeter, at uneven path depending on universities. Changes in accounting of capital expenditure. Potential difficulty of separation between education and healthcare in university hospitals.

Country	Coverage	Current expenditure: personnel	Current expenditure: non personnel	Capital expenditure	Breakdown: other	Remarks
Germany	HEIs: technical universities, technical colleges, theological colleges, colleges of education, colleges of art and music, universities of applied science, and colleges of public administration.	Available	Available			The university hospitals produce a very strong bias due to their large-scale medical services of a general hospital. Existence of two accounting systems with different level of aggregation (cameralistic and commercial).
Greece	Universities and TEIs	not individual institution Data is available	not individual institution Data is available	not individual institution Data is available		
Hungary	Higher education institutions (public only)	Available	Available	Available	By institution	Data sources: Ministry of Education and Culture
Ireland	Universities and Institutes of Technology	"Recurrent grants" in the Annual Accounts reports available. It may be possible to split between pay-related and non-pay-related.	"Recurrent grants" in the Annual Accounts reports available. It may be possible to split between pay-related and non-pay-related.	Available ("capital grants" in the in the Annual Accounts reports)	By institution	
Italy	University Higher education institutions	Personnel expenditure (permanent position personnel, contract personnel, other personnel, social contribution and other costs related to personnel).	Functioning costs (institutional expenditure, consumer goods and services, utilities, maintenance and locations) expenditure for students (grants, other expenditure) financial costs and taxes; debt reimbursement, transfers (for ordinary expenditure or investments), other functioning expenditure	Capital expenditure (estate expenditure, machinery, etc., financial assets), internal transfers (to departments, institutes, centres, etc.) for investments		Slightly different system for State and Non-State universities (less details for the latter). Data is generally available both according to accrual and cash criteria.

Country	Coverage	Current expenditure: personnel	Current expenditure: non personnel	Capital expenditure	Breakdown: other	Remarks
Latvia	Research active HEIs (15 public and 5 private)	Data (as 'salaries' plus 'compulsory payments of social insurance') available only for the 15 public HEIs	Data (as 'total expenditure' minus 'salaries' minus 'compulsory payments of social insurance' minus ' 'movable and immovable property') available only for the 15 public HEIs	Data (as 'movable and immovable property') available only for the 15 public HEIs	Salaries by academic and administrative staff; business trip expenses; payments for services; materials, energy resources; purchase of books and journals; scholarships, transport compensations; other expenses. Data available also specifically regarding R&D expenditure (total, current, salaries, capital).	All data on expenditure confidential from the information annually collected by the Central Statistical Office; Data on expenditure of private HEIs confidential from the information annually collected by the Ministry of Education and Science.
Lithuania	Public higher education institutions	Available	Available	Available	Breakdown by accounting categories not available	Breakdown available by the following categories: expenditure on studies, expenditure on RTD, expenditure on scholarships and other support, other direct expenditure (no further breakdown available), total personnel cost (brutto), personnel costs of teaching personnel, expenditure on purchased goods and services, expenditure for costs related to international exchange of staff, other expenditure (no further breakdown available), accrued interest on held accounts
Luxembourg	University of Luxemburg				By type of expenditure	
Malta	University of Malta	available - this includes both academic and non-academic personnel expenditure	available	available - this includes payments for capital expenditure obtained from cash flow statements		Educational expenditure of the University are available in the University's audited financial statements through the NSO. Latest data is available for the year 2008. In this respect, the UOM financial (academic year) covers the period 1st October 2007 - 30th September 2008. For National Accounts

Country	Coverage	Current expenditure: personnel	Current expenditure: non personnel	Capital expenditure	Breakdown: other	Remarks
						purposes, these accounts are treated as covering data for the whole 2008, since 9 out of 12 months are actually in 2008.
Netherlands	Public research universities and public universities of applied sciences. Problems relate to the (five) private universities.	Available	Available	Available (investments in buildings, equipment and other capital goods)	Current expenditure are broken down into: personnel exp, depreciation, expenditure on buildings/real estate (from 2008), and "other expenditure" (e.g. interest). No breakdown per educational field	Note: Expenditure (current expenses) are based on accrual accounting system. This means they include depreciation. There is a separate capital expenses account.
Norway	Public and private higher education institutions. Possibly som problems for defence institutions.	Available	Available	Has to be checked. Investments in buildings normally done outside regular budgets.	Breakdown by accounting categories available.	
Poland	Higher education institutions 5A	Available	Available	Available	By cost categories (depreciation, materials, energy, foreign services, taxes, personnel, insurance, R&D apparatus vs. running expenditure). By activity (education, R&D vs. running expenditure).	data on expenditure are available as % indexes only.
Portugal	Public Higher education institutions.	Available	Available	Available	By subunit (Faculties, Schools) when these have financial autonomy. By activity (education vs. R&D vs. other activities.	Costs of medicine are contested because of the difficult separation between education and healthcare in university hospital.
Romania	Higher education institutions	Available at aggregate level	Not available	Available at aggregate level		No breakdown for ISCED 5B institutions since they are part of ISCED 5A institutions.

Country	Coverage	Current expenditure: personnel	Current expenditure: non personnel	Capital expenditure	Breakdown: other	Remarks
Slovakia	Only for public Higher Education Institutions (HEIs)	available; however, only by selection of adequate data and their elaboration	available; however, only by selection of adequate data and their elaboration	available; however, only by selection of adequate data and their elaboration	By the categories which are subjects of the official statistical survey done for public HEIs	Data adequate to the specifications presented in the D4 are available only after the data selection from the financial database of public HEIs and their elaboration. Similar official statistical survey is not applied for private or state HEIs.
Slovenia	Higher education institutions (universities, single higher education institutions, vocational colleges)					Data on expenditure are collected on aggregated level. Expenditure for tertiary education also covers expenditure for research and development.
Spain	Public higher education institutions (47 public universities)	Available	Available	Available	By cost categories (following UOE manual), By activity (education vs. R&D)	Data not available by faculties or schools make it impossible to split the data into fields of study as UOE requested.
Sweden	Higher education institutions (universities, university colleges, university colleges of arts).	Available	Available	Available	By activity type (education vs. R&D activities).	No breakdown by subject domains
Switzerland	Data publicly available for the whole perimeter, with exclusion of the private UAS (2 small institutions).	available	available	not available since for many institutions still included in public budgets; there is an attempt to calculate full costs using standard rates per sqm of available space.	By educational field (based on a classification of institutional units into fields). By activity (education vs. R&D vs. other activities.	Slightly different system for universities and UAS (no full comparability possible). Cost of medicine are contested because of difficult separation between education and healthcare in university hospitals.
United Kingdom	Higher education institutions reporting to HESA which are involved in the Research Assessment Exercise (this restricted perimeter has been suggested by contact person at National Statistical Authority)	Available	Available		By institution	

Source: EUMIDA.

Research and development Expenditure

Classification in R&D Expenditure is unclear in some countries. The following overview presents information on methods used in ERA countries based on a recent EUROSTAT questionnaire, as well as information available in the Sources & Methods database of OECD.

Even if responses from some countries are not fully clear, evidence from Table 25 allows dividing the countries surveyed in four broad groups:

- Countries performing an annual R&D time-sheet survey of all personnel: these include Portugal, Romania, Sweden, Switzerland, and the UK.
- Countries performing regular surveys every few years and using them to derive coefficients for R&D activities for the following years. These include for example Finland, Ireland, and Norway.
- Countries using national coefficients, mostly based on some older surveys and which are not differentiated for individual institutions (but possibly for domain, personnel category and type of institution). These include Italy, France, Germany, and Greece.
- Countries using coefficients estimated by the statistical units themselves (e.g. heads of universities).

Usability of these methods for obtaining data on R&D expenditure of individual institutions has to be judged differently. Disregarding problems of quality of the survey and of self-estimates of use of time, data from the first group of countries can effectively be used at the institutional level; the same applies for the second group of countries if surveys are used to derive institution-specific coefficients, since it has to be assumed that coefficients don't change too rapidly from year to year.

Data from countries in the third groups cannot be used at the institutional level, since the coefficients are not specific for each institution. Finally, data from the last group of countries has to be used with much care since the quality of the coefficients might be very variable.

Table 25. Summary overview of availability of data on R&D Expenditure by country

Country	Method	Detail
Austria	Annual survey of personnel	R&D personnel expressed in FTE throughout. Comprehensive time-budget surveys are performed in all sectors and sub-sectors with the exception of the "Company R&D sub-sector": The following information was requested for each individual scientist and engineer and each individual technician employed in the units surveyed by time-budget surveys: i) Duration of employment in reference year and whether full-time, half-time or otherwise; ii) Actual total working hours per week (including overtime work) on average; iii) Estimate of percentage distribution of working hours among: -- administration -- teaching (only at universities) -- R&D -- other activities. On the basis of the detailed information provided by individual respondents, individual FTE data for each person were calculated by the surveying agency. No overall coefficients were used.
Belgium	Estimates at the level of statistical units (biannual)	The methods used to collect R&D data in the higher education sector were completely overhauled in 1995. Data is now collected by means of questionnaires sent to the central administrations of universities and university centres or institutes. In collecting data on funding resources and R&D personnel, priority is therefore given to data supplied by central administrations and other centralised sources of information, since the latter are considered to provide data that are more comprehensive and detailed than those collected separately from individual research departments or units. In addition, central administrations are felt to be better placed to provide estimates in cases where Data is missing.
Bulgaria	Estimates at the level of statistical units	Coefficients for estimation of R&D share of total HES Data is not used. Data for R&D variables on HES are compiled on the basis of information collected through a statistical survey on R&D activity of the units in this sector.
Cyprus		No coefficients are being used
Czech Republic	Estimates at the level of statistical units	No special R&D coefficients are used for the HES in the Czech Statistical Office. Individual faculties have different methods to estimate R&D expenditure and R&D personnel. Some faculties have even daily evidences of the percentage of time devoted to R&D.
Denmark		In the guidelines of the questionnaires it is stressed that no coefficients must be used. However, we are aware that at some universities FTE's and thus expenses for R&D is calculated according to the formal group contracts for different kind of positions, in which the expected share of R&D-work, teaching etc. is specified. The quality of the formal shares of R&D-work is expected to be low for the individual researcher, but more acceptable at higher levels of aggregation.

Country	Method	Detail
Estonia		No R&D coefficients are used at SE level. They are used by some large universities for the estimation of FTE data to report.
Finland	Coefficients from previous survey	The coefficients are used for the compilation of the universities' R&D. The coefficients (R&D shares) were computed from time use survey of the university teachers and researchers for the academic year 2004-2005. Time-use survey approximately every 10 years. The survey was similar to the survey for 1991-1992. The coefficients were computed by the main field of science and group of post. The results caused slight decline if the full-time equivalents of the researchers compared to results using the coefficients of the previous time use survey. However, the time use survey is extensive as every post is included for one week and the posts are distributed evenly over the whole calendar year.
France	Coefficients	University R&D resources are globally estimated for the whole of universities and institutes on the basis of the share of R&D applicable to different budget items. The rate of 50% is applied to personnel expenditure and to the calculation of FTE. This estimate is based on numerous data files supplemented by a survey on resources by university contract, a survey conducted by the departments in charge of research within the Ministry of Research. In 1997 the use of new administrative sources made it possible to produce a more accurate estimate of the number of lecturer/researchers, resulting in a downwards revision of the figures.
Germany	Coefficients	For certain groups of Grandes Écoles not administered by the Ministry responsible for higher education, in consideration of their large number, their R&D activities are from now on estimates based on a sample (approximately 1 out of 2).
Greece	Coefficients	The research coefficients were established through surveys in higher education units (Universities only, Technical Colleges are not yet included) and are used for the estimations on Labour costs & Capital expenditure on R&D as well as for the FTE. • Labour costs for R&D are estimated from total labour costs using different coefficients for different fields of science (mathematics and computer sciences 42%, physical sciences 39%, chemical sciences 43%, biological sciences 35%, earth and related (environmental sciences) 36%, engineering sciences 36%, medical sciences 34%, agricultural sciences 35%, social sciences 35%, humanities 38%). • Land and buildings: R&D expenditure are estimated to range between 35% and 47% of total capital expenditure on land and buildings (according to the field of Science as supplied by Ministry of National Economy). • Instruments and equipment: R&D expenditure are estimated to range between 36% and 51% of total capital expenditure on instruments and equipment according to the field of science (as supplied by Ministry of National Economy). • Other current costs are estimated to be 20% of total general university funds • Three groups of coefficients are used according to R&D personnel occupation. The first group refers to the teaching personnel, the second to administrative staff and technical, special teaching personnel, and the third group to support personnel. FTE for each member of supporting personnel is estimated to be 0.1 person year; it differs slightly according to specific labour duties. Postgraduate students FTE is estimated to be 70% of the total number. The calculations for the resources allocated to R&D are based on data and estimations: a) the Data is obtained directly (via a questionnaire) from the University Research Committees b) The estimations are made on data obtained from the Ministry of Education and the Ministry of National Economy. A system of coefficients, established through surveys in higher education units, is used in the estimations. The current coefficients were revised in 1995, previously in 1989, and a new revision completed and will be applied to the survey of 2005. The first survey for the setting of these coefficients was in 1983.

Country	Method	Detail
Hungary		We employ no such coefficients, as all the R&D expenditure and personnel data is aggregated from the data received from the organisational units themselves. Any estimation methods – should they be employed – in calculating respective R&D expenditure and personnel figures are to be found at the level of the individual respondents. Reporting units make the calculations of FTE for RSE, technicians and other personnel. For research institutes, calculations are made by heads of the smallest individual organisational units (sections, laboratories and workshops) based on the time spent on R&D. The same applies for R&D units of enterprises. In the higher education sector, R&D units only report weight ratios. Data on R&D are reported by staff members of the universities and colleges in an annual and mandatory survey. Data on R&D personnel are collected in headcount and in full-time equivalents and include persons employed by the institutions on a full-time or part-time basis. In the Higher Education Sector, R&D units report FTE in person-years for researchers, technicians and other personnel. Administrative personnel is included when supplying specific administrative support to R&D activities. Though these methods are different, Data is comparable from one sector to another.
Ireland	Coefficients from previous survey (2006/7)	A full time use survey is carried out every 4 years to allow for transition of data from HC to FTE and also to estimate the portion of the block grant dedicated to the pay of researchers. A higher quality time use model is being developed as part of the third level reform program - data systems will be operational by 2011. The last time use survey was conducted in 2006/7. a lot of effort was put into improving quality of data in this key part of the survey.
Italy	Coefficients	Full-time equivalent data on R&D are calculated as follows: researchers 100%; professors/lecturers 50%; technicians 54%; other personnel are calculated as a fixed 10% of the non-teaching personnel engaged in the three ancillary areas - general services, administrative services, and libraries. The higher education sector data relies on indirect estimates to determine the proportion of the expenditure and personnel devoted to R&D that is to be attributed to the sector as a whole. Such estimates are based on the use of coefficients to determine the proportion between teaching and research activities to be set against the various headings in higher education budgets and related also to the total personnel at institutes of higher education.
Latvia		
Lithuania		
Luxembourg		
Malta		
Netherlands	Survey	Beginning 1990 data reflect a change in methodology (implemented in 1994 but computed for the years 1990 onwards). R&D coefficients are calculated based on two sources: statistical data about total university personnel on December 31 and statistical data about the actual time spent on R&D by scientific personnel at the universities in a specific year. The new method results in a higher estimate of the R&D expenditure (on the average an increase in 20-25%) than the former one
Norway	Coefficients from previous survey (2000)	R&D coefficients obtained in a time-budget survey undertaken in 2000 for the universities; before the 2001 survey, coefficients based on a 1991 time-budget survey. R&D coefficients for the state university colleges are calculated from a

Country	Method	Detail
		time-budget survey from 1997. For 2005 coefficients are based on a time-budget survey conducted for 2005 for these institutions. For institutions not included in the time-budget surveys (some scientific university colleges) R&D coefficients are estimated in close collaboration with the institutions.
Poland		
Portugal	Survey	No coefficients are used to estimate R&D expenditure and R&D personnel in the HES sector. Each researcher answers about the time allocated to R&D activities in the unit (in the survey form to all researchers of the unit). To minimize the subjectivity of the answers (perceptions of the time) a method is applied. The method is based on the researcher answers (in terms of percentage of they're working time) and on a typology of periods of time committed to R&D: 10% (minimum), 30%, 50%, 70% and 100%. Consequently, the labour costs are estimated to. To accomplish that, the information about the average of public careers' salaries of the researchers and the new figures for the percentage of they're working time are used.
Romania	Survey	Due to the existent annual survey based on HES data collection, we did not use coefficients in order to estimate R&D expenditure and R&D personnel statistics.
Slovakia		
Slovenia		R&D coefficients are used for the calculation and estimation of FTE data to report in line with FM.
Spain		Until 2002, data on R&D activities in the higher education sector are in fact national estimates based on the accounting records of universities and the State Secretariat for Universities and Science. For the years in question, estimates were made of the share of general university funds (GUF) considered to be devoted to R&D, which was then added to resources allocated directly. The same type of adjustment was made with regard to personnel. The method of estimating the share of GUF actually devoted to R&D was developed by the National Statistics Institute in accordance with the methods set out in Annex_3 of the Frascati Manual (time-budget surveys and use of central administration data) and doesn't apply to public budget funding for R&D which are evaluated by another body on the basis of funding intentions that are not established by the same method.
Sweden	Survey	
Switzerland	Survey	The cantonal universities, federal institutes of technology, FIT research institutes and universities of applied sciences provide data regarding the number of working hours their research staff devotes to various activities, particularly R&D. Employee workweek percentage Data is supplied on an annual basis. In order to generate R&D expenditure statistics for the various types of tertiary-level institutions, working hours devoted to R&D activities are compared with financial and payroll statistics
United Kingdom	Survey	Data is obtained from HEFCE TRAC data which is collected from all the HEIs.

Source: EUMIDA

Funding sources

All institutions in Ireland, Spain, Sweden and UK present data on funding sources and there is also a good coverage of data from Cyprus, Germany, Italy, the Netherlands and Norway (Table 26).

As in the case of Expenditure, data on Funding sources has better coverage for public institutions. In Austria, Bulgaria, Czech Republic, Finland, Lithuania, Portugal, Slovakia and Spain this information is not available for private institutions. As for Expenditure indicators, in Cyprus this data is not available for private institutions but only at aggregated level.

Origin of funding sources is available in the majority of countries but classification in core budget, third-party funding and fees is not clear in some cases. In general, data on core budget includes information on state, regional or local budget (as in Czech Republic, France, Hungary, Romania, Spain and Switzerland). In some countries data on third-party funding is classified in public or private funding (Hungary, Italy, Romania, Slovenia, Spain, Sweden). Information about fees usually includes data on tuition fees by academic programs (not available in Bulgaria, Czech Republic, Ireland and Sweden). In Italy the breakdown between core and third-party funding may be problematic. Other breakdowns are: donations and grants by country or region (Table 27).

Table 26. Availability of data on funding sources by country (number of institutions without data)

	Missing cases	Not available	Not applicable	Confidential	Unreliable or uncertain data	Deviates from definition	Provisional value	Total no of institutions
AT				46	22			68
BE	1	12	2					43
BG	1	4		6				39
CH	1	4						36
CY		1			1			7
CZ		2						26
DE		5				301		306
EE				7				7
ES								47
FI		3						49
GR		40						40
HU	2	31						61
IE						21		21
IT		2						81
LT				16				16
LU								1
LV				5				20
MT				1				1
NL		1						59
NO								44
PL		91						91
PT		4						18
RO		57						57
SE								41
SI		1						4
SK		13						33
UK								148
Total	**2**	**271**	**2**	**81**	**1**	**322**		**1364**

Source: EUMIDA

Table 27. Summary overview of availability of data on funding sources by country

Country	Coverage	Core budget	Third-party funding	Fees	Breakdown: other	Remarks
Austria	Public universities	available but confidential, public provision currently discussed	available but confidential, public provision currently discussed	available but confidential, public provision currently discussed		
Belgium						Information will be checked
Bulgaria	35 accredited universities; BAS; ACA	available	Only donations are available. Research funding information is not available.	Not available	Data available by: Property income, Other non-tax incomes, VAT and other taxes, country donations, EU donations, donations from third countries	
Cyprus	Private and public higher education institutions (university and non-university level).	Available by Statistical Service at aggregate level for private institutions and disaggregated for public institutions.	Available by Statistical Service at aggregate level for private institutions and disaggregated for public institutions.	There are no fees for public institutions. Aggregate data available for private institutions.	Available by Statistical Service at aggregate level for private institutions and disaggregated for public institutions.	Data has been requested for 2008 for public Universities with the following breakdown: Government grant, Subsidy for undergraduate fees, Fees, Donations, other revenue. For Public Non-University same data breakdown exist as from before 2008. For Private University and Non-University data exist disaggregated from before 2008 broken down by: Main activity revenue and Other.
Czech Republic	Public HEIs of universities type	Dissagregation: revenues from sales, state subsidies	Available	Not available	Revenues from sales of products, goods services	
Denmark	Core university institutions: 8 universities and 6 university-like institutions	Available (government funding for core research and education)	Available (funding from research councils, government agencies, EU etc.)	Available (tuition)		The 8 universities produce a joint set of statistics. The indicated availability is based on this set. The remaining "university-like" institutions do not produce this kind of common figures, but there are chances that this may happen. Otherwise, data must be collected by contacting each institution.

Country	Coverage	Core budget	Third-party funding	Fees	Breakdown: other	Remarks
Estonia	Public and private universities			Fees from Academic programs (ISCED 5A and 6) and Professional programs (ISCED 5B)		Other available data: Public funding (state, local government), Private funding (households, legal persons), Institutions' own funding, International sources
Finland	Universities and Polytechnics in the field of Ministry of Education and FI031	Universities: data on expenditure by source: budgetary and external funding (i.e. third party). External sources by Academy of Finland, Tekes, Domestic company, other domestic, EU, Foreign company, other foreign financing. Polytechnics: basic state funding, supplementary funding from state, special funding, finance from owner (municipal, federation of municipalities, private)	Universities: External funding. Polytechnics: Fees for services (usually research contracts)	No fees for tuition. Polytechnics: Income from services for fee (usually research contracts) considered as third party funding	Expenditure data from universities generally available by performance area (educ, research) and education fields (not fully compatible with ISCED97)	As universities act as state agencies, it is sometimes impossible to make difference between expenditure and revenues. No data is likely recorded from FI031. Generally: breakdown core/third party in not like a problem.
France	All institutions of the extended data set	Data on: State permanent funding (from MESR, from other ministries), Public subsidies and grants (from local authorities, EU, ANR, other)	Information on Research contracts is available	Core resources (students fees, PRI revenues, long life learning fees, taxe d'apprentissage (a), other resources from services provided, other)		Potential difficulty of separation between education and healthcare in university hospitals
Germany	HEIs: technical universities, technical colleges, theological colleges, colleges of education, colleges of art and music, universities of applied science, and colleges of public administration.	Data disaggregated by administrative revenues, and current basic funding	Available	student fees available		Third-party funding is readily available, the rest needs to be edited by FSO

Country	Coverage	Core budget	Third-party funding	Fees	Breakdown: other	Remarks
Greece	Universities and TEIs	not individual institution Data is available	not individual institution Data is available. Only R&D funding by individual university is available. R&D funding for TEIs Data is not considered reliable	not applicable	R&D funding by university is broken down by source as follows: central government, Decentralised government services, Own resources, Public enterprises, Private enterprises, Other foreign sources European Union-research framework programme, Community Structural Funds, Private non-profit organisations, Private foreign enterprises, Foreign governments, Foreign higher education institutions, Foreign private non-profit organisations, International organisations	
Hungary	Higher Education Institutions (universities and colleges)	Data on Government funding, EU and international funding, donations	Data available: public and private contract			
Ireland	Universities, Institutes of Technology and colleges that are funded by HEA	Data on funding received from HEA is available, split into Recurrent grants, Capital grants, Erasmus grants, Research grants.	data collected by Forfas but not released due to confidentiality agreements undertaken with universities.	Data on tuition fees is not available	Other available data: Recurrent grants, Capital grants, Erasmus grants, Research grants	
Italy	University Higher education institutions	available	available	available		Breakdown between core and third-party funding may be problematic
Latvia	Research active HEIs (15 public and 5 private)	Data (as 'endowment from general income') available only for the 15 public HEIs	Data (as 'total funding' minus 'endowment from general income' minus 'student fees') available only for the 15 public HEIs	Data (as 'student fees') available only for the 15 public HEIs	Funding from the EU structural funds; research funding; international funding for studies; income from the rent of premises; other income. Data available also specifically regarding R&D funding (total, state budget, funding from international sources, research funding by the HEI,	All data on revenues confidential for the information annually collected by the Central Statistical Office; Data on revenues of private HEIs confidential for the information annually collected by the Ministry of Education and Science.

Country	Coverage	Core budget	Third-party funding	Fees	Breakdown: other	Remarks
					other).	
Lithuania	Public higher education institutions	Available (assignations from the state budget)	Available	Available	Breakdown available by: public budget assignations used for scholarships and other support, international exchanges, capital expenditure, total revenues received from legal and physical persons, revenues from legal and physical persons used for scholarships, total revenues received from foreign institutions, international organisations, funds, programmes, projects (revenues received directly, of which amount used for scholarships, revenues received via state-governed institutions, of which amount used for scholarships, revenues received from student fees, other revenues from students, revenues received from foreign institutions, international organisations, funds, programmes, projects ffor provided services	
Luxembourg	University of Luxembourg	will be checked	Will be checked			

Country	Coverage	Core budget	Third-party funding	Fees	Breakdown: other	Remarks
Malta	University of Malta	available - this includes the budget given to University by government and interests from investments	available - this includes funds from donations, funds from EU Institutions and Programmes and other types of funds. Fees from companies for educational services are included in the fees section and not in third-party funding.	available - this includes tuituion fees and other fees charged for educational services and other services. Fees paid for other welfare services furnished to students by the educational institutions is not applicable.		Revenues of the University are available in the University's audited financial statements through the NSO. Latest data is available for the year 2008. In this respect, the UOM financial (academic year) covers the period 1st October 2007 - 30th September 2008. For National Accounts purposes, these accounts are treated as covering data for the whole 2008, since 9 out of 12 months are actually in 2008.
Netherlands	Public research universities and public universities of applied sciences. No data for (five) private universities	Only available: Core funding (recurrent funding) from the government	available (defined as: third party funding)	Available: private fee revenues	Breakdown also has Other Revenues (e.g. from interest), next to Core budget, fees and Third party funds	It is not possible to get data on what parts of revenues were provided for education and what for R&D
Norway	Public and private higher education institutions. Possibly som problems for defence institutions.	Available	Available	Fees not in operation for public institutions. Available for private institutions.	Detailed breakdowns by accounting reports available.	
Poland	Higher education institutions 5A	available	available	Distinction by public and private fees		
Portugal	Public Higher education institutions.	Available (Public HEIs)	Available (Public HEIs)	Available (Public HEIs)	Data disaggregated by sub-unit (Faculties, Schools) when these have financial autonomy and by types of sources (governments transfers, services)	
Romania	Higher education institutions	Disaggregated in: state budget, local budget data available at aggregate level	available: private funds and other/third-party funds data available at aggregate level	Classified in: tuition fees, donations, sponsorships, scholarships, funds from NGOs, funds from international organizations, credits awarded to students by economic agents, NGOs or other private entities data available at		It is not possible to get data on what parts of this was provided for education and what for R&D

Country	Coverage	Core budget	Third-party funding	Fees	Breakdown: other	Remarks
				aggregate level		
Slovakia	Only for public HEIs	available; however, only by selection of adequate data and their elaboration	available; however, only by selection of adequate data and their elaboration	available; however, only by selection of adequate data and their elaboration	By the categories which are subjects of the official statistical survey done for public HEIs	Data adequate to the specifications are available only after the data selection from the financial database of public HEIs and their elaboration. Similar official statistical survey is not applied for private or state HEIs.
Slovenia	Higher education institutions (universities, single higher education institutions, vocational colleges)	Data on revenues can be disaggregated by 3 categories: public funds (from the state, local governments and from EU), private funds (from households and private companies for education) and international funds (non EU).	Data on revenues can be disaggregated by 3 categories: public funds (from the state, local governments and from EU), private funds (from households and private companies for education) and international funds (non EU).	Data on revenues can be disaggregated by 3 categories: public funds (from the state, local governments and from EU), private funds (from households and private companies for education) and international funds (non EU).		Data is at the moment available only for year 2007. For 2008 data will be prepared at the end of June.
Spain	Public higher education institutions (47 public universities)	Included: Patrimonial funds, alienation of real investments	Disaggregated in Other public institutions funds, European Union funding, private funds	Data available by students fees, national and regional government funding		
Sweden	Public and private higher education institutions	Available	Available	Not-applicable	Breakdown by source and level of education	
Switzerland	Available for the whole perimeter expect the two private universities of applied sciences	Division between core budget and third-party funds already in the statistical data (based on detailed list of funding items).	Available, with a high degree of disaggregation by sources (types of funding programs, public vs. private, etc.)	Available		

Country	Coverage	Core budget	Third-party funding	Fees	Breakdown: other	Remarks
United Kingdom	Higher education institutions reporting to HESA which are involved in the Research Assessment Exercise (this restricted perimeter has been suggested by contact person at NS Authority)	Available: includes Funding body grants	Available: includes research grants & contracts	Available: includes tuition fees & education contracts	Information about Joint ventures, Endowment & investment income, non-research income for services rendered to industrial and commercial companies and public corporations	

Source: EUMIDA

Personnel

Data on personnel is available in the majority of cases, only not available in Portugal and in a fey number of institutions in Belgium (Table 28). Disaggregation in personnel categories is frequent in all countries, which includes classification in academic and non-academic personnel and with more detail by professional category. Breakdown by field is available in general except in Cyprus, Czech Republic, Ireland, Malta, Slovakia, Slovenia and Spain. Usually data is collected in FTE but in Greece, Italy, Luxembourg, and Poland only in headcount. In Malta, data is available in headcounts but is also broken down between full-time and part-time personnel. Data on PhD students is heterogeneous because there are countries in which this personnel is not included as academic staff (Cyprus, Czech Republic, Greece, Ireland, Italy, Latvia and UK) and in other countries this information is not available. In the majority of cases there is information on personnel by gender, age and nationality (Table 29).

Table 28. Availability of data on personnel by country (N institutions without data)

	Missing cases	Not available	Not applicable	Confidential	Unreliable or uncertain data	Deviates from definition	Provisional value	Total no of institutions
AT	34	1				33		68
BE	1	8						43
BG	5	5			4			39
CH		1						36
CY								7
CZ		2						26
DE								306
EE		2				1		7
ES								47
FI		3						49
GR	40							40
HU								61
IE								21
IT		1			1		79	81
LT				1				16
LU								1
LV						20		20
MT	1							1
NL								59
NO								44
PL								91
PT		18						18
RO				1				57
SE								41
SI						4		4
SK	13							33
UK								148
Total	**94**	**41**	**0**	**2**	**5**	**58**	**79**	**1364**

Source: EUMIDA

Table 29. Summary overview of availability of data on Personnel by country

Country	Coverage	Personnel categories	Breakdown by field	FTE	PhD Students	Remarks
Austria	Public universities, Universities of applied science, Private universities, University colleges of teacher education	different according to type of HEI; total staff not available for AUS	available	Data is available at FTEs and headcounts	third-party financed PhD students included in staff	Data could be obtained also by gender
Belgium						Information will be checked
Bulgaria	All Higher education institutions recognized by MEYS - through ADMIN data base	Only academic staff	Available	only as headcounts, ADMIN does not consider FTE	Available	NSI has data in all variables but considers it confidential
Cyprus	Public and private higher education institutions (university and non-university level).	Number of teaching (by: category), non-teaching and management personnel are available (disaggregated data available for private institutions after permission given for EUMIDA project). For the Non-University level data is broken down only by: Teaching and Headmasters. The non-teaching personnel breakdown applies to all institutions the same as above.	Not available by field of science	Data is available at FTEs and headcounts	PhD students are not included in personnel figures, but in student figures. (Disaggregated Data is available for public institutions. For private institutions Data is available after their permission for EUMIDA project).	The number of R&D personnel and researchers of business enterprises, government, higher education and private non-profit sector, are available by occupation, qualification, gender
Czech Republic	Public HEIs of university type	Research personnel, professors, assistant professors, assistants, lecturers	Not available	All Data is available in FTE: academic and other, research staff, professors, assistant professors, lecturer etc. Data in headcounts are available only for whole HEIs, but could be obtained from annual report of HEIs	not included in staff	data on headcount are available by gender

Country	Coverage	Personnel categories	Breakdown by field	FTE	PhD Students	Remarks
				in details.		
Denmark	Core university institutions: 8 universities and 6 university-like institutions	Faculty (full-time and part time);	Faculty broken down on research and teaching	FTE is the general measure for all personnel data	Salaried PhD students may be included; must be checked	The 8 universities produce a joint set of statistics. The indicated availability is based on this set. The remaining "university-like" institutions do not produce this kind of common figures, but there are chances that this may happen. Otherwise, data must be collected by contacting each institution.
Estonia	Higher education institutions (universities, professional higher education institutions, vocational education schools)				will be checked	Breakdown by quite many dimensions should be possible (age, degree, gender, educational domain). At the moment, there is no good source of data. Possible sources are as follows. -The wage survey conducted by the Statistics Estonia included all entities with at least 50 employees. -The register of the instructors of higher education institutions as part of the Estonian Educational Information System will start in the Fall 2009. The first data collection was made but it is still not possible to use that data because the educational institutions are still correcting the submitted data
Finland	Universities and Polytechnics in the field of Ministry of Education and FI031	Universities: Budget funded teaching (Profs and assoc. Profs, Senior assistants, Assistants,	Universities: ISCED97 not likely available, but good approximation using domestic Fields of	Universities: FTEs as required. Polytechnics: headcounts in cross section.	Universities: available Polytechnics: NA	Polytechnics data refer only partly to whole calendar year (tenure teachers as cross section)

Country	Coverage	Personnel categories	Breakdown by field	FTE	PhD Students	Remarks
		Lecturers, Full-time teachers, Full-time equivalent hourly teaching) + externally financed teaching. Research staff (2 groups + other staff: Polytechnics: Teachers(tenure), part time teachers, outsourced teaching, supporting staff in numerous categories.	education. Research personnel available according FOS2007 Polytechnics: not available according ISCED97			
France	All institutions of the extended data set	Data on academics and non academics staff	Data on specialization (SISE-CNU fields, not readily compatible with OUE)	Both headcounts and FTEs	Possible marginal problems for PhD and post-doc position	Data available by gender, age and nationality (academic staff)
Germany	HEIs: technical universities, technical colleges, theological colleges, colleges of education, colleges of art and music, universities of applied science, and colleges of public administration.	Data available by occupational description (professors, lecturer, scientific assistant, scientific and artistic employees, teaching and administrative staffs)	available (national classification will be transformed into UOE fields of education by FSO)	Data is both in FTEs and headcounts	available	Desegregation by gender, age and nationality exists but not all data is publicly available due to confidentiality issues
Greece	Universities, TEIs	Available (for academic and administrative staff)	available only for academic staff	Data is in headcounts	PhD students are not included in personnel figures, but in student figures.	Data available by gender. Academic staff Data is also available by rank.
Hungary	Higher Education Institutions (universities and colleges)	Data disaggregated by: academic staff (teachers), technical and administrative staff (non-teachers).	data available by faculties	Available	Available	Data available by gender and age
Ireland	Universities, Institutes of Technology and colleges that are funded by HEA	Disaggregated by academic and non-academic, full time and part time, permanent and temporary	data not available by field	available	PhD students are not considered part of university staff; they are enclosed in the student record	
Italy	University Higher education institutions	Data disaggregated by academic staff (permanent, non permanent) and non	Available by field of science	Only head count is available	Not included in personnel figures, but available	

Country	Coverage	Personnel categories	Breakdown by field	FTE	PhD Students	Remarks
		academic (permanent and non permanent).				
Latvia	Research active HEIs (15 public and 5 private)	Data available on personnel (broken down by academic and general; gender) in main work	Data available for academic staff (broken down by field of science) of the 15 public HEIs	Data available in both FTEs and headcounts	PhD students in general are not included in the staff data unless they are individually employed by the HEI.	Further breakdown of data on academic staff by professional category, age group, workload is available; Data on the breakdown of R&D staff by field of science, collected by the Central Statistical Bureau, is confidential but publicly available from the Ministry of Education and Science. Breakdown of academic staff by national/foreign available only for visiting academic staff.
Lithuania	Public and some private higher education institutions. University hospitals not covered.	Available by detailed breakdown on type of positions.	Available only for RTD personnel according to Frascati	Available by full-time and part-time	Available	
Luxembourg	University of Luxembourg	Data by type of personnel, distinguishing between academic and non-academic personnel, as well as more detailed categories	will be checked	Data is in headcounts	will be checked	Data available by nationality, gender and age
Malta	University of Malta	Data is available for academic and non-academic personnel	not available	Data is available in headcounts and broken down between full-time and part-time	Data includes academic posts occupied/ taken up within the University	Data for the academic year 2007/2008 is available from NCHE. Data is also available by gender, age, qualification level, grade and salary scale. Breakdown of academic staff between national and foreign staff is not available.
Netherlands	Public research universities and public universities of applied sciences. University hospitals not fully covered	Type of personnel (scientific versus support staff)	For research universities breakdown available for each of the eight disciplinary domains. For Universities of Applied Sciences data not broken down.	Data is both in FTEs and headcounts.	Research universities personnel includes data on PhD students employed by these institutions	Data available by gender and age. Research university personnel in medical disciplines are not covered fully by data collection since they are employed by medical centres (academic hospitals) that operate quite separately from

Country	Coverage	Personnel categories	Breakdown by field	FTE	PhD Students	Remarks
						their university.
Norway	Public and private higher education institutions. Possibly some problems for defence institutions.	Available by detailed breakdown on type of positions.	Available	Available by headcount and FTE.	Available.	
Poland	Higher education institutions 5A	Data by type of personnel (academic, non-academic as well as detailed break-down by personnel categories of academics and non-academics)	will be checked	Headcounts but not FTEs	There is no systematic data on employed PhD students	Data available by gender and nationality. Data on staff are available as % indexes only. Some data on staff are available under the restriction (for the official use)
Portugal	Public Higher education institutions.	Available: professional situation and academic category	Available by scientific area	both headcounts and FTEs (FTE exist but we are still waiting to know if they will be made available by the ministry)	will be checked (not included as part of staff, unless they are hired as teaching staff. the two issues are independent)	Data available by gender, age, place of residence and nationality
Romania	Higher education institutions	disaggregated in: academic, non-academic, administrative, maintenance etc	not available	All personnel in education are counted (full and part-time personnel)	PhD students are not accounted as academic staff	Data available by gender and groups of age. Data broken by nationality and field of education or field of science are not available.
Slovakia	Higher education institutions; however, for private and state HEIs the survey is limited only on the categories dealing with the academic staff.	Statistical categories: full professors, associated, researchers, other academic staff, technical and administrative staff	this is not a subject for the official statistical survey	available	this is not a subject for the official statistical survey	Non-academic staff is not a subject for the official statistical survey on level of the private and state HEIs. The statistical survey of the academic staff is available for all HEIs and for example there are categories as age, gender so on, but not nationality.
Slovenia	Higher education institutions (universities, single HE institutions).	data available distinguishing between academic and non-academic personnel	not available	Data is both in FTEs (only for academic personnel) and headcounts	data available for so called "young researchers" (PhD students in special governmentally financed program)	Data available on degree of education and gender

Country	Coverage	Personnel categories	Breakdown by field	FTE	PhD Students	Remarks
Spain	Public higher education institutions (47 public universities)	Disaggregated by: Type of personnel (academic vs non-academic staff) and professional category (full professor, associate, assistant, other position)	not available	Both head counts and FTEs	PhD students are not accounted as academic staff	Data available also by age and gender
Sweden	Public and private higher education institutions	Available by detailed breakdown on type of positions.	Available	Available by headcount and FTE	Available	
Switzerland	Available for the whole perimeter (including private UAS). All data in FTE (Headcounts also available).	There is a fairly detailed breakdown in personnel categories, whose application is however not always very uniform across institution. Breakdown between administrative and technical personnel available.	Available based on a detailed breakdown of fields which is compatible with UOE.	Available	There is no separate PhD students category in the personnel statistics; it is known that most PhD students are in fact employed by the university itself, but no matching of data possible.	
United Kingdom	Higher education institutions reporting to HESA which are involved in the Research Assessment Exercise (this restricted perimeter has been suggested by contact person at National Statistical Authority)	Distinction is between academic and non-academic staff and by academic employment function (teaching-only, teaching & research, research-only, neither teaching nor research). Data on Grade (academic staff only) (professors, senior lecturers, lecturers, researchers, others) are also available	available by academic cost centre. Concordance between these categories and OECD FOS classification needs to be made.	available	PhD students are not considered part of university staff; they are enclosed in the student record	

Source: EUMIDA

Educational activities

As shown by Table 30, information on students is available in all countries with a high coverage (only for public institutions in the Czech Republic) and in general FTE are considered (in Finland only headcount data is available). Disaggregation by program level is available in all cases but there is no information about adult education and distance education. In all cases, information is also available by nationality, age and gender (see Table 31).

Table 30. Availability of data on students (ISCED 5) by country (number of institutions without data)

	Missing cases	Not available	Not applicable	Confidential	Unreliable or uncertain data	Deviates from definition	Provisional value	Total no of institutions
AT		1						68
BE								43
BG			6					39
CH								36
CY								7
CZ								26
DE		6						306
EE								7
ES								47
FI		1				48		49
GR			2			37		40
HU								61
IE						21		21
IT			2		3			81
LT				1				16
LU								1
LV								20
MT								1
NL	1							59
NO								44
PL								91
PT								18
RO				1				57
SE								41
SI								4
SK								33
UK								148
Total	**1**	**8**	**10**	**2**	**3**	**106**	**0**	**1364**

Table 31. Summary overview of availability of data on Educational Activities by country

Country	Coverage	Headcount	Breakdown by field of education	Program level	Remarks
Austria	Public universities, Universities of applied science, Private universities, University colleges of teacher education	available	available by field of study	available	
Belgium					data available by nationality, gender and year
Bulgaria	All Higher education institutions recognized by MEYS - through ADMIN data base	available	available fields of education and specialty studied	available	data available by citizenship, year and mode of attendance. NSI considers it confidential
Cyprus	Public and private higher education institutions (university and non-university level)	available (disaggregated data available for private institutions after permission given for EUMIDA project)	available (disaggregated data available for private institutions after permission given for EUMIDA project)	available (disaggregated data available for private institutions after permission given for EUMIDA project)	Data available by age, nationality, gender, community/religious group, country of citizenship, duration of study and level of study. Data on graduates are also available.
Czech Republic	Public HEIs of university type	available	not available	not available	data available by nationality, gender and year
Denmark	Core university institutions: 8 universities and 6 university-like institutions	Annual enrolment, total enrolment, degrees awarded	Four main categories: The humanities, the social sciences; the natural sciences and engineering, and medicine/health	Three main categories: Bachelor, MSc, and Diplomas/MBAs etc	The 8 universities produce a joint set of statistics. The indicated availability is based on this set. The remaining "university-like" institutions do not produce this kind of common figures, but there are chances that this may happen. Otherwise, data must be collected by contacting each institution.
Estonia	Higher education institutions (universities, professional higher education institutions, vocational education schools)	available	available: Field of education (up to the 3-digit ISCED 97 code)	available	data available by nationality, gender and year
Finland	Universities and Polytechnics in the field of Ministry of Education and FI031	Headcounts and FTEs available	available	available	data about foreign students, international mobility, total students by gender are also available

Country	Coverage	Headcount	Breakdown by field of education	Program level	Remarks
France	all core set institutions		Data available: Specialization (SISE-CNU classes or ad-hoc classes, not directly compatible with OUE)	available	higher risk of some double counting is detected since there is no unified codification of the student in the whole higher education system. Possibility to obtain data on age, gender and nationality
Germany	HEIs: technical universities, technical colleges, theological colleges, colleges of education, colleges of art and music, universities of applied science, and colleges of public administration.	available	available (national classification will be transformed into UOE fields of education by FSO)	available	Other available data: share of students in the 1st semester and share of graduates in age-specific population, ratio of students to teaching staff, enrolment rates, distribution of graduates by fields of education, share of international students, distribution of international students by home country and field of education, distribution of graduates by gender.
Greece	Universities and TEIs	available	available	available (only universities offer degrees at ISCED 6)	Other available data: enrolled students by gender, nationality, age. Data for graduates can also be broken down by gender and age.
Hungary	Higher education institutions (public only)	Available	Available	Available	By institution
Ireland	Universities, Institutes of Technology and colleges that are funded by HEA	data available	available	available	Data available: Enrolments, new entrants and graduates by institution, mode of study (full time/part time), gender, field of study and qualification (undergraduate and postgraduate)
Italy	University Higher education institutions	data available	will be checked	available according to Bologna process. Unique cycle (5-6 years) still exist for some undergraduate courses also in new order curricula.	Other available breakdown: gender, nationality, age, career, etc.
Latvia	Research active HEIs (15 public and 4 private)	Data available both on students and degrees	Data available both on students and degrees	Data available both on students and degrees	Headcounts. Breakdown by gender and citizenship

Country	Coverage	Headcount	Breakdown by field of education	Program level	Remarks
					(national/international) students/graduates available.
Lithuania	Public and private higher education institutions.	Available	Available according to both State Code of Educational Programmes and ISCED Code at ISCED 5 level and for ISCED 6 - data on research field according to CERIF	Available	Data on the students and degrees available also for international students
Luxembourg	University of Luxembourg	data available	will be checked	available	Data available also by gender and age
Malta	University of Malta	Data is in headcounts	available	available	For the academic year 2006/2007, data for both enrolled students at ISCED 5 and 6 levels and graduations at ISCED 5 level is available from NSO. For the academic year 2007/2008, data is only available for enrolled students at ISCED 5 and 6 level from NCHE.
Netherlands	Public research universities and public universities of applied sciences	Available	available by Field of education	available	Data available by gender
Norway	Public and private higher education institutions.	Available, except defence institutions only in the aggregate for all institutions.	Available according to national standard. Can be converted to OEU standard.	Available	Data on students and degrees available by national/international.
Poland	Higher education institutions 5A	will be checked	available by Field of study	available	Data available by nationality, gender, age, e-learning studies
Portugal	Public Higher education institutions.	available	available by educational domain	available	Data available by gender, age and origin. Bologna transition, problems of consistency in the treatment of data
Romania	Higher education institutions	available	available	available	Data available by gender, age, citizenship, foreigners, by counties
Slovakia	All higher education institutions	available	available; however only by mapping between the national classification and UOE (ISCED97) classification	available	Data available by gender, age, nationality, year-class etc.

Country	Coverage	Headcount	Breakdown by field of education	Program level	Remarks
Slovenia	All public and private universities, faculties, academies of art and professional higher education institutions executing accreditated HE-programs by Council for Higher education of the Republic of Slovenia	data available	data available	available	Data available by gender and on aggregated level also by age, origin, foreigners
Spain	Public higher education institutions (47 public universities)	available	available by field of study	available	Data available by Gender, age, origin, nationality
Sweden	Public and private higher education institutions	Available	Available	Available	
Switzerland	Data available for the whole perimeter.	available	available	Division between diplomas, bachelor, master available; in Switzerland there are still a few programs at master-level without a bachelor, but these could be rather easily identified since they are concentrated in a few sectors (e.g. medicine)	
United Kingdom	Higher education institutions reporting to HESA which are involved in the Research Assessment Exercise (this restricted perimeter has been suggested by contact person at National Statistical Authority)	data available	Available by subject of study (subjects classified using JACS codes; concordance between JACS and fields of education as in the UOE manual needs to be made)	available	Postdoctoral students are not included. Information about all students by institution, gender, ethnicity and disability status is available

Source: EUMIDA

Research and technology production

Number of ISCED6 degrees

Table 32 shows that there is a good coverage for this indicator. Total degree awarded is only not available for Bulgarian higher education institutions.

This information is available for both public and private institutions (only in Czech Republic, Netherland, Spain and Switzerland there is information for public HEIs only). The number of ISCED6 degrees is available in all countries (except in Bulgaria) and the same data is disaggregated by gender, nationality, and age (see Table 33).

Table 32. Availability of data on total degrees awarded by country (number of institutions without data)

	Missing cases	Not available	Not applicable	Confidential	Unreliable or uncertain data	Deviates from definition	Provisional value	Total no of institutions
AT		1	4					68
BE		4				32		43
BG		39						39
CH		2						36
CY			1					7
CZ								26
DE		8						306
EE								7
ES								47
FI		1				48		49
GR		1	2			37		40
HU		1						61
IE								21
IT					6			81
LT				1				16
LU								1
LV								20
MT								1
NL		2						59
NO								44
PL								91
PT								18
RO				1				57
SE								41
SI								4
SK								33
UK								148
Total	**0**	**59**	**7**	**2**	**6**	**117**	**0**	**1364**

Table 33. Summary overview of availability of data on Research and Technology production by country

Country	Coverage	Number of ISCED6 degree	Publications	Patents	Remarks
Austria	Public universities, Private universities	Number of PhD students are available			
Belgium	information will be checked	information will be checked			
Bulgaria	none	not available	not available	not available	
Cyprus	Public and private higher education institutions (university level ISCED 5A, ISCED 6)	Number of ISCED6 degree available by field of education and nationality			
Czech Republic	Public HEIS of university type	Available			
Denmark	Core university institutions: 8 universities and 6 university-like institutions	Available	Research publications are counted by the Ministry	University-owned patents are counted by the Ministry	The source of data on PhD enrolment and degrees is Statistics Denmark
Estonia					Breakdown by: Educational domain, age, type of degree, language of study, source of funding
Finland	Universities under field of ministry of education	Number of PhD students are available.		Number of patents where university or polytechnic is first applicant also available.	

Country	Coverage	Number of ISCED6 degree	Publications	Patents	Remarks
France	All core set institutions	Number of PhD degrees awarded is available			
Germany	HEIs: technical universities, technical colleges, theological colleges, colleges of education, colleges of art and music, universities of applied science, and colleges of public administration.	Available (number of awarded degrees)			
Greece	Universities only	Number of ISCED6 degree available by field of education (not by nationality)			
Hungary	Higher Education Institutions (universities and colleges)	Available			
Ireland	Universities, Institutes of Technology and colleges that are funded by HEA	Data available: degrees obtained by gender, institution & field of study, divided according to type of course			
Italy	University Higher education institutions	Available. Breakdown by nationality is available; breakdown by field may be difficult because of cross field programs.			
Latvia	Research active HEIs (15 public and 5 private)	Data available by gender, year of birth, field of education, citizenship (national/ foreign)			
Lithuania	Public and private higher education institutions having the right to issue a doctorate	Available			

Country	Coverage	Number of ISCED6 degree	Publications	Patents	Remarks
Luxembourg	University of Luxembourg	Data available also by level of education, gender and age			
Malta	University of Malta	Available			Latest data on graduations at ISCED 6 level is available from NSO for the academic year 2006/2007.
Netherlands	Public research universities (have the right to award PhD)	Data on PhDs available by gender and field of education			
Norway	Public and private higher education institutions having the right to issue a doctorate	Available			
Poland	Higher education institutions 5A	Data available by nationality, gender, age, e-learning studies and field of education			
Portugal	Public Higher education institutions.	Available			
Romania	Higher education institutions	Data available by gender, age, field of education, foreigner, counties			
Slovakia	All higher education institutions with the accreditation for the PhD study	Data available by gender, age, nationality and field of study (national classification - there is possibility for mapping into the UOE / ISCED67 classification).			
Slovenia	All public and private universities, faculties, academies of art and professional higher education institutions executing accredited HE-programs by Council for Higher education	Data available on aggregated level by field of education, gender and age.			

Country	Coverage	Number of ISCED6 degree	Publications	Patents	Remarks
	of the Republic of Slovenia				
Spain	Public higher education institutions (47 public universities)	Data available by Gender, age, origin, nationality and field of study	ISI publications		
Sweden	Public and private higher education institutions having the right to issue a doctorate	Available			
Switzerland	Public universities only (other institutions don't have the right of awarding PhDs	Available, including breakdown by field of education			
United Kingdom	Higher education institutions reporting to HESA which are involved in the Research Assessment Exercise (this restricted perimeter has been suggested by contact person at National Statistical Authority)	Available			

Source: EUMIDA

Additional information

Considering additional data, Table 34 demonstrates that information on R&D expenditure is unavailable for 5 countries and confidential in a higher number of institutions in other 5 countries. Total funding from private sector is unavailable for 14 countries; private funding for R&D is unavailable for 9 countries and data on patents and spin-offs is unavailable for 16 or more countries.

Table 34. Availability of data on R&D expenditure by country (number of institutions without data)

	Missing cases	Not available	Not applicable	Confidential	Unreliable or uncertain data	Deviates from definition	Provisional value	Total no of institutions
AT				46				68
BE	3	33						43
BG	3	5		3				39
CH	1	16						36
CY								7
CZ		26						26
DE		306						306
EE				7				7
ES								47
FI		3						49
GR	1	17	2					40
HU		4						61
IE						21		21
IT		81						81
LT				16				16
LU							1	1
LV				5				20
MT				1				1
NL		59						59
NO				35				44
PL		91						91
PT								18
RO				57				57
SE								41
SI								4
SK		13						33
UK								148
Total	**8**	**654**	**2**	**170**	**0**	**21**	**1**	**1364**

Source: EUMIDA

Scientific publications

In most country it is possible to obtain information on publications in Web of Science but there are some problems with access, normalization and institutional count of documents. In some countries there are institutional databases and annual reports containing information of publications in all documental types (for a share of HEIs at least). Some problems are detected in Bulgaria, Cyprus, Germany, Greece, Italy, Latvia and Malta.

Patents

As for publication indicators, information on patents is problematic because usually there are not official source to obtain this data. National Patent Office in the EUMIDA countries or European Patent Office could provide some data but normalization and aggregation at institutional level is not clear in most countries. Some problems are detected in Austria, Bulgaria, Finland, Germany, Greece, Hungary, Ireland, Italy, Latvia, Luxembourg, Norway, Portugal, Romania, Slovakia and Switzerland.

Technology transfers activities

It is difficult to obtain this indicator because in the majority of countries no sources and no systematic data collection exists. Some problems are detected in Austria, Bulgaria, Cyprus, Germany, Greece, Hungary, Italy, Latvia, Luxembourg, Malta, Norway, Portugal, Romania and Sweden.

Employability

In some countries data on employability is provided by the HE institutions which produce annual surveys to analyze the activity of their graduates or from the Ministry of Labour, but there are different levels of aggregation and difficulties with comparability between countries. Some problems are detected in Austria, France, Germany, Greece, Hungary, Latvia, Luxembourg, Poland, Romania and Slovenia.

4.4.2 Reasons for non-availability

For the extended set of data, reasons for non-availability are detected in general with data related to expenditure and incomes. Some problems concerning availability are detected in Austria, Cyprus, Estonia, France, Germany, Greece, Latvia, Lithuania, Malta, Romania, Slovakia and Slovenia. For data concerning personnel, students and graduates, the principal problems are the aggregation level (in Bulgaria, Portugal, Romania and Slovenia) and the reasons for non-availability for private institutions (in Cyprus). In Ireland, data on personnel is not public but could be accessible upon request. Data on students and graduates is not available only in Slovenia and Romania.

Interaction with National Experts suggested that the reasons for non-availability could be classified as follows:

(a) Legal issues (e.g. the statistical law explicitly forbids the publication of microdata)

(b) Administrative barriers (e.g. it would be possible to publish microdata but NSI depend on administrative decisions from the Ministry)

(c) Institutional settings (e.g. there is no obstacle, but the publication of microdata requires additional workload for which the institution is not prepared, or it would be difficult to allocate the responsibility internally).

4.4.3 Summary overview

Table 35 displays a summary overview of the state of data by country. Mainly problems are detected with the availability of data related to outputs and with reasons for non-availability of data on expenditure and incomes.

Table 35. Summary overview of availability and reasons for non-availability by country

Country	Availability	Reasons for non availability
	Some difficulties in	**Some difficulties in**
Austria	Total staff (1 inst) Staff by field (46 inst) Total N students (1 inst) ISCED 5 by field (1 inst) Total degree awarded (1 inst) Degree by field (1 inst)	Total Expenditure (46 inst) Incomes (46 inst) R&D expenditure (46 inst)
Belgium	Total expenditure (13 inst) Incomes (12 inst) Total staff (8 inst) Staff by field (37 inst) ISCED 5 by field (43 inst) ISCED 6 by field (17 inst) Total degree awarded (4 inst) Degree by field (43 inst) R&D expenditure (33 inst)	
Bulgaria	Total expenditure (4 inst) Incomes (4 inst) Total staff (5 inst) Staff by field (5 inst) ISCED 6 (5 inst) ISCED 6 by field (5 inst) Total degree awarded (39 inst) Degree by field (39 inst) R&D expenditure (5 inst)	Total expenditure (6 inst) Incomes (6 inst) R&D expenditure (3 inst)
Cyprus	Total expenditure (1 inst) Incomes (1 inst) Staff by field (7 inst)	
Czech Republic	Total expenditure (2 inst) Incomes (2 inst) Total staff (2 inst) Staff by field (26 inst) ISCED 5 by field (26 inst) ISCED 6 by field (26 inst) Degree by field (26 inst) R&D expenditure (26 inst)	
Denmark		
Estonia	Total staff (2 inst) Staff by field (4 inst)	Total Expenditure (7 inst) Incomes (7 inst) R&D expenditure (7 inst)

Country	Availability	Reasons for non availability
Finland	Total expenditure (3 inst) Incomes (3 inst) Total staff (3 inst) Staff by field (1 inst) Total N students (1 inst) ISCED 5 by field (1 inst) ISCED 6 (1 inst) ISCED 6 by field (1 inst) Total degree awarded (1 inst) Degree by field (1 inst) R&D expenditure (3 inst)	
France		
Germany	Total expenditure (5 inst) Incomes (5 inst) Total N students (6 inst) ISCED 5 by field (6 inst) Total degree awarded (8 inst) Degree by field (8 inst) R&D expenditure (306 inst)	
Greece	Total expenditure (40 inst) Incomes (40 inst) ISCED 5 by field (1 inst) ISCED 6 (1 inst) ISCED 6 by field (1 inst) Total degree awarded (1 inst) Degree by field (1 inst) R&D expenditure (17 inst)	
Hungary	Total Expenditure (32 inst) Incomes (31 inst) Staff by field (61 inst) Total degree awarded (1 inst) Degree by field (61 inst) R&D expenditure (4 inst)	
Ireland	Staff by field (21 inst)	
Italy	Total expenditure (2 inst) Incomes (2 inst) Total staff (1 inst) Staff by field (1 inst) ISCED 6 by field (80 inst) R&D expenditure (81 inst)	
Latvia		Total Expenditure (5 inst) Incomes (5 inst) Staff by field (5 inst) Total N students (1 inst) ISCED 5 by field (1 inst) R&D expenditure (5 inst)
Lithuania	Staff by field (16 inst)	Total Expenditure (16 inst) Incomes (16 inst) Total staff (1 inst) ISCED 6 (1 inst) ISCED 6 by field (1 inst) Total degree awarded (1 inst) Degree by field (1 inst) R&D expenditure (16 inst)
Luxembourg	Staff by field (1 inst) ISCED 5 by field (1 inst) ISCED 6 by field (1 inst) Degree by field (1 inst)	
Malta	Staff by field (1 inst) R&D expenditure (1 inst)	Incomes (1 inst)

Country	Availability	Reasons for non availability
Netherlands	Total expenditure (1 inst) Incomes (1 inst) Staff by field (59 inst) ISCED 6 (1 inst) ISCED 6 by field (2 inst) Total degree awarded (2 inst) Degree by field (2 inst) R&D expenditure (59 inst)	
Norway		R&D expenditure (35 inst)
Poland	Total expenditure (91 inst) Incomes (91 inst) Staff by field (91 inst) ISCED 6 by field (91 inst) R&D expenditure (91 inst)	
Portugal	Total expenditure (4 inst) Incomes (4 inst) Total staff (18 inst) Staff by field (18 inst)	
Romania	Total expenditure (4 inst) Incomes (57 inst) Staff by field (57 inst)	Total Expenditure (1 inst) Total staff (1 inst) Total N students (1 inst) ISCED 5 by field (1 inst) ISCED 6 (1 inst) ISCED 6 by field (1 inst) Total degree awarded (1 inst) Degree by field (1 inst) R&D expenditure (57 inst)
Slovakia	Total expenditure (13 inst) Incomes (13 inst) Staff by field (33 inst) R&D expenditure (13 inst)	
Slovenia	Total expenditure (1 inst) Incomes (1 inst) Staff by field (4 inst)	Total Expenditure (2 inst)
Spain		
Sweden		
Switzerland	Total Expenditure (5 inst) Incomes (4 inst) Total staff (1 inst) Staff by field (1 inst) Total degree awarded (2 inst) Degree by field (2 inst) R&D Expenditure (16 inst)	
United Kingdom	ISCED 6 by field (2 inst)	

Source: EUMIDA

4.5 Actors and roles in data collection

In general, the Ministry of Education and the National Statistic Office are the main owners of data. In most countries there are official sources that provide most of indicators for the core set and extended set of data. In the case of "descriptors" (*year of foundation*, *region*, etc.), data can obtained directly from each institution.

In the following, a summary of ownership by country is presented.

Austria

For the core dataset the Austrian federal statistical office has information on *students, staff, university hospital, specialization, highest degree* while the Ministry of Research and Science can provide information about *legal status*. There are not official sources to obtain information on *control* because it is more difficult to distinguish government dependent public and private (some of the private universities receive strong support from federal governments).
Most data considered in the extended set is collected and managed by the Austrian federal statistical office. The Federal Ministry for Science and Research (FMSR) has however also established a detailed database on higher education statistics, uni:data. Statistics on educational *expenditure* are annually published by Statistik Austria on aggregate. Self-reported data on *scientific publications* and *patents* is included in the Intellectual Capital Report of public universities. No systematic data collection on *technology transfer activities* exists, however the Austrian Research and Technology Reports (published yearly by the FMSR) cover aspects of technology transfer activities of public universities irregularly. Data on labour market entrance (*employability*) and early careers of university graduates from younger cohort is not available. For some study programmes distinct reports exist, mainly initiated by particular universities/faculties themselves.

Belgium

In this country there are diverse institutions providing indicators on Higher Education. Data on higher education is provided by the "Observatoire de l'enseignement supèrieur" for the French Community, "Administratie Onderwijs en Vorming, afdeling Hoger Onderwijs" for the Flemish Community and the "Ministerium der Deutschsprachigen Gemeinschaft, Abteilung Unterricht, Ausbildung und Beschäftigung" for the German-speaking Community. The Frenchspeaking and Flemish offices publish, each for its own community, indicators relative to *highest degree*, *students, graduates* and *doctorates*. Data on scientific research can also be obtained from Belspo, the federal office for Belgian science policy.
In Flanders, statistics on scientific *publications, patents* and statistics on *spin-offs* are collected every year per university, as part of the allocation of research funds (BOF) to the university. Sources and availability of data for *expenditure*, *funding sources* and *employability* have to be cross-checked for availability.

Bulgaria

The majority of data is available from the National Statistical Office: *country, region, control, higher education, student, graduates, university hospital, research activity* and *specialization*. Most of this data is confidential however. Ministry of Education and Science can provide information about *specialisation, highest degree, students* and *staff* (for a

number of HEIs). No official statistical sources is detected for *foundation year* but this data is generally available through the tax register BULSTAD or the marketing register DAKSI.
Information on *expenditure* and *funding sources* is available through the Ministry of Education and the Ministry of Finances (for 51 accredited universities). Data on *personnel*, *students*, and *graduated* could be collected for National Statistical Office but only in aggregate format. There are no official sources to obtain information on output (*scientific publications, patents*) and *transfer activities*.

Cyprus

In Cyprus, the Ministry of Education and Culture owns the information about *country, year, highest degree, doctorates, educational expenditure*, *revenues* and *students* (only for public institutions, private are confidential), while the Statistical Office provides data for *control, enrolled students, specialization* and Information on *personnel*, *graduates* and *employability* (only for public institutions). Information on *research activities* is available from HEIs. There are no official sources to obtain information on *publications* and *transference activities* but each institution publishes periodically a list of their teaching staff's publications. Data on applications for *patents* is available by the Department of Registrar of Companies and Official Receiver of the Ministry of Trade, Industry and Tourism.

Czech Republic

Most of the data is collected, elaborated and published by Institute for Information on Education (IIE, Department of Statistical Information and Analysis) that is established by Ministry of Education, Youth and Sports (MEYS) and is the main authority in statistical services on education. Czech Statistical Office (CSO) and its Education, Health, Culture and Social Security Statistics Section respectively is responsible for life-long learning only. Some information about institutions is provided by MEYS itself in annual report on HEIs. In the case of *specialization,* a transformation of existing data is need.
Data about *expenditure* and *revenues* of public HEIs is based on reporting to the MEYS in the annual reports on economic management of individual institutions. Data about *personnel* are annually collected by IIE from public universities and published in Statistical Yearbook on Education. *Student's* information could be obtained from MEYS. Data on *scientific publication* and *patents* are available in the database called "Information Register of R&D results" (RIV), elaborated by R&D Council. There is no specific survey about technology transfer activities of Czech HEIs. Data on unemployment of graduates is collected by the labour offices, gathered by the Ministry of Labour and Social Affairs and published by the Centre for Educational Policy (at Charles University, Prague) in a web-based database.

Denmark

According to the Denmark National Report, all indicators for EUMIDA project could be collected from Statistic Denmark that develop the main data-base for educational purposes so-called "student register". On the university sector specifically, the association of Danish Universities collects more detailed data (capital structure, funding, publications).
Statistics Denmark collects data on all public *research expenditure*, especially covering university research. For the traditional Universities - governed by the Ministry of Science, Technology and Innovation (MSTI) - the association of Danish Universities collects data on *funding sources*, costs, *personnel*, capital structure, buildings/premises, *students*, *research publications*, and some rudimentary numbers on collaboration/interaction with society at large. Regarding research output, MSTI is introducing a standard classification and registration system, much inspired by the Norwegian bibliometric system. MSTI entertains a register of university patents. It covers university-owned patents only.

Estonia

The Ministry of Education and Research (EHIS) collect the statistical data of educational institutions offering higher education; in addition, the Statistical Office of Estonia is also collecting the data in aggregate form (available since 1980, but meanwhile the whole structure of educational institutions and diplomas has been changed). All information to get indicators for the core set of data could be obtained from the Ministry of Education and Research.
The Statistical Office regularly collects data on higher education expenditure in private and public universities directly from the involved institutions. Data on *students* and *graduates* are available from the Ministry of Education and Research (Estonian Education Information System) and data on *personnel* will be collected after 2010. At the moment this information could be obtained from each institution. It should be possible to produce the statistics on publications using the Estonian Science Information System (but at individual level not aggregated for institution). There is no systematic data on *patents* collected for Estonian HEIs. Some data is published in annual reports of universities and Estonian Patent Office publishes aggregate data. There is no data concerning *technology transfer* activities regularly collected at the moment.

Finland

Statistics Finland executes the nationwide data collection on all the polytechnics and universities for the production of official statistics and international data needs (Eurostat, OECD, UNESCO). Statistics Finland also maintains the original individual level database collected from universities' and polytechnics' study registers. According with information presented in National Report, all indicators could be obtained from Statistics Office. There

are not official sources for *foundation year, university hospital* and *distance educations* but this information is easily available from each institution.
Data on *expenditure* is available for the majority of universities from the Ministry of Education (KOTA). Statistics Finland's contain data on R&D *personnel, students* and graduated for universities in the field of Ministry of Education. This Ministry, also records data on FTE research years and the number of academic *publications*. No public statistical sources give complete information on university *patents*. The most reliable solution would be to ask universities themselves.

France

There is no official, comprehensive publication of data or statistics on HEIs at micro-level, neither on education (to the exception of total number of students) nor on research and third mission. The lowest level of official data aggregation the most commonly used is the "académie", an administrative level at which all education and part of the research mission are piloted. The national statistical office (INSEE) only publishes a very limited number of official aggregated data on the universities. All is produced by other entities: the Ministry for Higher Education and Research (MESR) collects data on all institutions under its supervision, and sometimes mixes this data to some other data about institutions under the supervision of other ministries. Some other Ministries (e.g. Agriculture, Defence, Industry, Culture etc.) are also gathering data about the HEIs under their supervision, but this data is usually not published. Apart from the MESR, the most relevant institution in France for data on academia is the OST (Observatory of Sciences and Technologies). Indicators for the core set of data could be obtained, in general, from the MESR.
Information on *expenditure* and *funding sources* could be collected from the MESR (and possibly the Ministry of Budget). No individual data on expenditure is available on a public base. *Expected income* is available on line since Sep. 2009 for universities only. Data on *personnel* is owned by the MESR. Data is public; some is readily available on-line. Fairly detailed data on *students* and *degrees* of ISCED 5A and ISCED 6 institutions is gathered and published by the MESR, including at individual level. Data is public and available on-line for the extended data set (all engineering schools), except for business schools. There has not been so far any official data on *publications* and *patents* in the French Higher education System at institutional level. However, the OST is collecting data from standard sources (Thomson Reuters Scientific) and sends the results to each institution to check it. The same is made by OST with *patents* applicants from INPI (French patent office) and EPO. For data on *technology transference* the main source is the survey made by BETA. The CEREQ organisation is in charge of producing indicators about professional insertion of young people 3 years after the award of their diploma (initial training only, i.e. excluding students in long life education) that could be used as *employability* indicator.

Germany

Most of the data considered in the German National Report is collected and managed by the Federal Statistical Office (FSO), which is in charge of preparing the data delivery for international higher education and R&D statistics. However, for ISCED 5B, primary Data is to a large extent managed by the *Länder* themselves. Some data is available on-line on the FSO website. From FSO could be collected data for *country, region, students, doctorates* and *specialization*. There are no official sources for *foundation year*, but this information can be obtained from each institution.

Data on HEI *expenditure* is provided by FSO and annually extracted from the administration reports of HEIs. The national Statistical Office uses a so-called "R&D factor of core funding" that approximates the expenditure for fields of education. The FSO provides data on *funding* sources of HEIs (ISCED 5A/6), *personnel*, *students* and *graduates*. In Germany, *publications* and *patents* of universities are not recorded on a regular basis. A source for publication data is a report of the "*Centrum für Hochschulentwicklung*" (CHE) which provides rankings of German universities based on a variety of indicators including publications and citations. There is no systematic data about the respective *technology transfer* offices (TTOs) of universities or *employability*.

Greece

There are, currently, different data sources for Higher Education institutions. The National Statistical Service (NSS) collects data on Higher Education institutions including Universities and Higher Technological Institutes. The Ministry of Education also collects data on Higher Education Institutions, through questionnaires addressed annually to individual Higher Education Institutions. Information about *students* and *doctorates* could be collected from these two sources and NSS could provide data on *specialization*. There are official sources where to collect *foundation year* information but it is possible to obtain this directly from each institution.

The Ministry of Education does collect *expenditure* on higher education institutions, but only aggregate data for the higher education sector, mainly through the General Accounting Office. The General Secretariat for Research & Technology collects data on funding sources. There are two sources of data on *personnel* disaggregated at the level of individual institution: the surveys of the National Statistical Service and the surveys of the Ministry of Education. Detailed data on *students* and *graduates* is readily available for Higher Education Institutions (Universities and TEIs) and third level professional schools. No data on *scientific publications* is available for Higher Education institutions. No data on *patents* is collected directly from higher education institutions. Data on patents is available by the National Patent Office and includes higher education institutions. Significant limitations apply, as it is possible that higher education institutions also apply directly to the European Patent Office. Data on number of spin-offs of Higher Education institutions is available through a recent

one-off survey conducted by the General Secretariat for Research and Technology. No data is available on careers and *employability*.

Hungary

According to the comments presented in the National Report, data is available from different sources (institutional register and data collections). Besides that, due to the transformation of higher education institutions, the relevant Hungarian statistical system is under revision. The Central Statistical Office (HCSO) and the Statistical Unit of the Ministry of Culture and Education is responsible for collecting data from the HEIs (on faculty level). In this line, core indicators on *control, students, doctorates, university hospital* and *specialization* and *highest degree* could be obtained from the Ministry, while the Central Statistical Office could provide data for *research activities*. Basic institutional information (*country, foundation year, region*) could be collected from the register (Ministry of Education and Culture) or directly from each HEI.

Data on *research activity* of HEIs is collected by HCSO. The data on *expenditure* and *incomes, personnel, students and graduates* is collected by the Ministry of Culture and Education from all HEIs. The Hungarian Academy of Sciences (HAS) is a forerunner to make it compulsory for their members to provide a full list of *publications* on the HAS's website. The scientific *publications* in this database are based on the Web of Science and its relevant databank but it contains also other publications that are not covered by those databanks. There is no systematic data collection on *patents* of HEIs. Data relating to spin-off policy, number of spin-off companies and employability is not available for Hungarian HEIs.

Ireland

The sources of data for the core set of Irish institutions are twofold: HEA (Higher Education Authority) and Evidence Ltd. Most data about the Irish higher education sector is collected and disseminated by the Higher Education Authority. While HEA is the statutory planning and policy development body for higher education and research in Ireland. In addition, it is the funding authority for the universities, institutes of technology and a number of designated higher education institutions. Data for *foundation year* and *region* could be provided directly from each HEI.

Some information about the *research activities*, *funding sources* and *expenditure* of higher education institutions has been collected by Forfás (Ireland's national policy advisory body for enterprise and science) through their Higher Education Research and Development Survey. Higher Education Authority collects data on academic *staff* in higher education institutions. The units responsible for this collection are the Recurrent Grants & Institutes of Technology Sections. Data on *students* and *graduates* are collected from HEA. Data on *publications* by Irish universities in Web of Sciences are available from Evidence Ltd. There

are no official sources to obtain data on *patents*. Numerous data on technology transference carried out by Irish higher education institutions are being collected by the Department of Enterprise, Trade and Employment. Destinations of graduate statistics are produced by HEA, and are collected in a report that is published on a bi-annual basis.

Italy

The Italian Ministry of Higher Education owns the information needed to obtain indicators on *country, region*, *highest degree, control* (only data about state and non state institutions)*, students, doctorate,* and *specialization*. There are not official sources for data on *foundation year* and university hospitals, but this information is easily collected directly from each institution.

Data on HEI *expenditure* is regularly collected in the dataset on university balance sheets, which is managed by the Ministry of Universities and Research (MIUR). Also data on *funding sources* are collected by MIUR. Data on *personnel* could be obtained from MIUR-CINECA and the MIUR's Statistical Office. The latter institution could provide data on *students* and *graduates*. There is no systematic data collection on *patents* and *technology transfer activities* of universities, done by official sources. Nevertheless, a periodical survey is carried out by the NETVAIL among the universities belonging to the Consortium. Data includes information on patenting TTO and liaison offices but is not complete for the whole universe of HEIs. Data on careers and *employability* is collected by ISTAT through a survey on the transition from study to labour market that is regularly taken every third year.

Latvia

Most of the data considered in the national report on both higher education and tertiary vocational education is collected and managed by the Central Statistical Office of Latvia (CSB) (Division of culture, education, science and health statistics), which is in charge of preparing the data delivery for international higher education and R&D statistics. In this line, CSB could provide information on *highest degree, students, doctorates* and *research activity* while data on *country, control, university hospital* and *research activity* could be obtained from the Ministry of Education and Sciences.

CSB and the Ministry of Education and Science regularly collect data on the *expenditure, funding sources, personnel, students* and *graduates* of individual HEIs directly from the involved institutions. Data confidentiality applies to the data collected by CSB on expenditure, incomes and R&D personnel (by field of science) for all HEIs and to the same data collected by the MOE for private HEIs. Data on the Web of Science *publications* of Latvian institutions on an irregular basis have been produced and analysed by the Centre for Science and Technology Studies at the Latvian Academy of Sciences. More detailed reports have been produced for internal use of each university, but they are not publicly available.

There is no systematic data available on both national and international *patents* filed by Latvian universities and other HEIs. Data on the applications and registrations of national patents and some international Data is collected and managed by the Patent Office of the Republic of Latvia. *Technology transfer activities* are not annually and/or coherently accounted for neither by the CSB nor the Ministry of Education and Science. There is limited data on the *employment* situation of HEI graduates, yet some studies have been conducted.

Lithuania

All the EUMIDA-relevant data on HEIs are collected annually by Statistics Lithuania, according to UOE Manual. Some data is also collected and managed by the Ministry of Education and Science and made publicly available at micro level through the web interface called the Science Potential of Lithuania. However there are most other sources that provide information with different level of aggregation, therefore data must be collected from a diverse number of owners. National Statistical Office provides information on *students*, *doctorates* and *highest degree*, HEIS about *research activity*, the Research & Studies Monitoring and Analysis Centre for students, the Registry Centre for *foundation year*, and the Ministry of Education and Science for *control*. No official source to obtain data for *university hospital* is detected but this data is easily available from the individual Hospitals Annual Reports.

Data on *expenditure*, *funding sources, publications and technology transference* on both public ISCED 5A and 5B institutions is uploaded on-line yearly by individual institutions to a database maintained by Institute of Mathematics and Informatics. Some of the collected data is not available to public, including data on expenditure and incomes. Ministry of Education and Sciences provide data on *personnel* and Department of Statistics to the Government of the Republic of Lithuania provide data on *students* and *graduates*. There is no systematic data on *patents* and on *employability* but data by individual institutions is in some cases available in the annual reports some of the HEIs make available to public.

Luxembourg

This country includes only one public university, therefore data from HE could be obtained directly from the university or the National Statistic Office. This institution has data about *country, highest degree, student, doctorate, research activity* and *specialization*. There are not statistical sources for *year* but this is easily obtainable from the university.

The University of Luxembourg publishes on its web site every year (since 2006) its annual reports, which contain detailed information as regards *expenditure* on educational costs. Some information as regards *funding sources* is again available in the University annual reports. Fairly detailed data on *personnel, students* and *graduates* is collected yearly. Some detailed information for each research unit as regards *scientific publications* is provided also

in the annual report. Additional detailed information is available in the ISI database. Detailed information is available in *patent* databases (PATSTAT, EPO, USPTO). The availability and level of disaggregation of this information and data on *technology transfer activity* and *employability* still needs to be further investigated.

Malta

The limited number of public and state funded educational institutions makes data collection easier. In Malta, the National Statistics Office has been requesting data from private and non-state funded institutions through the Malta Statistics and Authority Act for a number of years. However, data is not complete. Recent efforts, mainly through the National Commission for Higher Education (NCHE), are providing a more complete data set. Indicators on *highest degree, student* and *doctorates* could be provides from NSO and the other could be collected directly from each institution.

Educational *expenditure* and *funding sources* of the University of Malta are available in the University's audited financial statements through the NSO. Data on academic and non-academic *personnel* of the University of Malta is readily available. Data on *student* population and *graduates* is also available. There is no systematic data on scientific *publications* and *technology transfer activities*. Data on *patents* is also not readily available through the NSO. Data on the *careers of graduates* is available for University of Malta graduates through a regular survey to a number of graduates, which results are published and can be accessed online.

Netherlands

Data on higher education is collected and managed mainly by CBS, the Dutch national statistical office Statistics Netherlands and by the Ministry of Education, Culture and Science. In general, indicators for the core set of data could be obtained from these institutions, while information about students and doctorates could be collected from the Association of Research Universities. Descriptors are obtainable consulting directly each institution.

For public universities data on *expenditure* and *funding sources* for individual institutions are regularly collected by the Central Funding of Institutions Agency and published in their annual report Financiële GegevensWetenschappelijk Onderwijs. For private universities data will have to be collected from the institutions themselves, based on their annual financial reports. For the research universities a breakdown between educational and research activities is not available. For the public universities, detailed data on *students* and *degrees* is collected yearly through the 1cHO collection. This database is maintained by the representative organisation for the UAS institutions (the HBO-raad), the Association of research universities (VSNU), the Education Ministry, the Statistical Agency CBS, the IBG agency, and the CFI agency. Data on general *publications* is collected by the universities

themselves and stored in the national KUOZ database maintained by the VSNU. Data on WoS publications for individual universities is collected by the CWTS group and published by the Netherlands Observatory of Science and Technology. There is no systematic data on *patents* filed by Dutch universities. *Technology transfer activities* are a very heterogeneous set of efforts, structures and results. There is data on graduate careers and graduate *employability* resulting from the yearly survey carried out by the Research Centre for Education and the Labour Market.

Norway

The data considered for EUMIDA project is collected and managed by three different providers: Database for Statistics on Higher Education (DBH), which contains data reported by institutions on most aspects of HEI activity. This source is owned by the Ministry of Education and Research, and operated by Norwegian Social Science Data Services. The second provider is Statistics Norway, which is in charge of preparing statistics for the higher education sector for international use (Eurostat), excluding R&D statistics. The third is the Norwegian Institute for Studies in Innovation, Research and Education (NIFU STEP) that is in charge of preparing the national R&D statistics for the higher education sector and provide also R&D statistics for international use (Eurostat and OECD). Data for *year of establishment* and *organisational changes* has to be obtained from institutions.
Data on public HEI *expenditure*, *funding sources*, *students*, *graduates* and *personnel* is regularly reported directly from the involved institutions and published for each institution. Data could be collected through Ministry of Education and Research / Database for Statistics on Higher Education. Also data on share of R&D activities could be obtained from NIFU STEP. For ISCED 5A institution Norway has for the purpose of indicator-based funding a system reporting all types of peer reviewed *publications*. There are not official sources to obtain information on *patents*. There is no systematic and permanent system for collecting data on *technology transfer activities* in Norway. There is fairly good data on the *employment* situation of graduates from higher education institutions collected through a regular survey performed biannually by NIFU STEP.

Poland

Most of the data is collected and managed by the Central Statistical Office (CSO) - branches of Gdansk and Szczecin. The Information Processing Centre in Warsaw collects some. Data on education is managed by the Central Statistical Office (Gdansk branch). For a few years the Central Statistical Office Statistical Office in Gdansk has been responsible for education data - that could be provides information on *students, doctorates* and *highest degree*- and Statistical Office in Szczecin is responsible for R&D data. There are no official sources for *foundation year* but this information is easily obtainable from the web site of each

university. There are no objections to presenting the data referring to *university hospital* but some problems at the point of contact (university and university hospital) have been detected by the National Expert.
Data on HEIs (5A) *expenditure* is regularly collected by CSO directly from the institutions but published annually in aggregation form. The situation concerning *funding sources* is similar to expenditure. Detailed data on *personnel, students and graduated* of ISCED 5A institutions is collected yearly by the CSO. Data referring to *publications*, *patents* and *technology transfer activities* on the level of the individual university organizational units (Faculties, Departments, etc.) will be available from The Information Processing Centre (IPC), on the basis of the previous year. There is no systematic data referring to the careers and *employability* in Poland. CSO intends to start to collect it in near future.

Portugal

Data is reasonably available for the public universities and public polytechnics. Data on the private sector is far more limited, being mostly restricted to data on *students* and *graduates*. Information about *students* could be obtained from the Ministry of Science, Technology and Higher Education. The other indicators (data on year, region and other characterization variables) have been collected by the National Expert.
Data on *expenditure* and *funding sources* is available only for public HEIs. A detailed account of their expenditure can be obtained through the Annual State's Financial Accounts. The Ministry of Higher Education publishes aggregated data on *academic staff* however, this data is not made available per institution, The other two sources are non public: the Ministry's of Higher Education and Science Database on Academic Staff and the Ministry's of Higher Education and Science Database on Human resources. Data on *students* and *graduates* is available and most is published on the web. In recent years, the ministry of HE and Science has been making a significant effort to develop a Bibliometric database based on WOS *publications*. This data is presented for the whole Portuguese scientific system and not per institution. There is no systematic data on *patents* filed and *technology transference activities* by Portuguese universities. Considering *employability*, since 2008, the Ministry of HE and Science has been publishing a report each semester, based on data on the number of Graduates unemployed and registered in Employment Services.

Romania

In this country, almost all indicators of the data set about higher education are obtained and validated by the National Institute of Statistics (NIS) and the results obtained are provided to Ministry of Education, Research and Innovation, International Organisations and EUROSTAT. NIS provides information on *total staff*, *number of students*, *specialisation in subject domains*, *highest degree delivered*, *number of doctorates awarded* for each public

and private university, but this is publicly available only in aggregate format. It could be made available by unit only after individual agreement of the providers.
Data on *expenditure*, *personnel*, *students* and *graduates* are collected at individual level by NIS. However, only the aggregate data is available to the public, while the individual data is confidential. CNFIS (National Council for financing of higher education) provided a study in 2007 related to some sort of classifications of public universities regarding research and including the number of scientific *publications* broken down into ISI articles, national articles, and scientific books. There is no data collection of the number of *patents* for ISCED 5A in Romania. Data on *Technology transfer* and *Employability* is not collected consistently.

Slovakia

The Ministry of Education of the Slovak Republic (MS SR) is the principal owner of all data that were used in the framework of the EUMIDA project. The database of this data and surveys is administrated by the Institute of Information and Prognosis of Education (UIPS) and the data collection to both EUMIDA surveys was provided by the UIPS and partly by the Ministry of Education of the Slovak Republic as well. All national and UOE statistics, dealing with "Education" and presented by the Statistical Office of the Slovak Republic, are ensured and elaborated by the UIPS as well.
For *expenditure* and *funding sources* indicators, aggregated values for the HE in Slovakia are available from the Slovstat database (SO SR) or unique (specific) date for HEIs are available from the State Treasury (Data is applied for UOE statistics). UIPS provides data on *personnel* at aggregated level. Fairly detailed data on *students* and *degrees* of HEIs are collected yearly by UIPS. There are fairly good values on the *employment* situation of graduates from higher education institutions collected through a regular survey performed by the Centre of Labour, Social Affairs and Family (guaranteed by the Ministry of Labour, Social Affairs and Family of the Slovak Republic) two times a year. ISI publications Category of the International Scientific Index (SCI) is elaborated by help the database operated in the framework of the Central Register of the Evidence of Publication activities in Slovakia. There is no systematic data on *patents* filed by Slovakian HE Institutions. Nowadays, partial information about patents is visible in the "Annual report" of the HEI oriented on results of education and RTD. Information about "Spin – off" companies in the form of aggregated statistics by sectors is partially accessible on the centralised information portal for research, development and innovation (CIP VVI). Information about revenues from the collaborative RTD projects realised in the framework of entrepreneurial activities is available from the Annual financial report of the HEI.

Slovenia

The Slovenian Office of Statistics collects data on Higher Education sector from the middle of the '80s providing indicators on *institution category*, *region*, *research activity*, *university hospital*, *distance education*, *specialization*, *highest degree*, *doctorates, students* and *staff*. Data can be provided for single institutions only after an agreement with individual HEIs.
Data on *expenditure* for formal education, *incomes*, *personnel, students* and *graduates* is regularly collected by Statistical Office of the Republic of Slovenia (SORS) but is only available at aggregate level without written authorisation of HEIs. Data on *expenditure for R&D* is a part of the Research and development, science and technology section in Statistical Yearbook of the Republic of Slovenia 2008. To measure R&D *publication* output and *patents* at HE-institutions in Slovenia, it is possible to use the data produced by Institute of Information Science (IZUM) in Maribor. The Slovenian intellectual property office also provides the data about the number of patents applied for at the European Patent Office (EPO), United State Trade and Patent Office (USTPO) and (separately) at Slovenian Patent Office. Data about other types of technological output at HEIs in Slovenia is not available at SURS. This type of data is possible to obtain only on the ground of direct observation of HE institutions. The data on the *employment* situation of graduates from higher education institutions is not collected at SORS.

Spain

The National Institute of Statistics (INE) and the Ministry of Education (MEC) publish annual data on the education sector. Both, the Council of University Coordination (CCU) and the National Institute of Statistics report systematic data, collected for each university institution. The information gathered includes basic data on the university including university name, *region, foundation year*, scientific fields covered, *specialisation* (general vs. technical), whether there is a *university hospital*; *students* data and data on other human resources. This information is available for both public and private universities.
Expenditure data and *incomes* are collected from the biannual publication "La Universidad Española en Cifras" (Spanish Universities' Figures) published by CRUE (Vice-Chancellors Conference of the Spanish Universities). Information on *personnel* is provided by the National Institute of Statistics (INE, several years). Enrolment data for *students* and *graduates* is provided by the annual publication of the Council of University Coordination from the "Estadística Universitaria" (university statistics) from 1990s the latest being for 2008. *Publication* data could be collected from the Web of Science Database and number of *patents* applied could be obtained from the "Oficina Española de Patentes y Marcas" – OEMP and the European Patent Office (EPO). It is also possible, via negotiation with "Fundación CyD" (the Universities Knowledge and Development Foundation), to obtain information from each university's technology transfer office. This information can be gathered from the

technology transfer offices of the Spanish public universities. *Employability* data could be obtained from research by the Council of University Coordination and ANECA.

Sweden

The Swedish National Agency for Higher Education (HSV) is the main and responsible national public agency for the development and collection of statistics on HEIs. HSV is responsible for the collection of various statistics on undergraduate and graduate education as well as on university financing. Other agencies providing statistics on some aspects on HEIs are Statistics Sweden (collects statistics on R&D of firms and organizations as well various aspects on personnel of employees in higher education institutions) and Swedish Agency for Higher Education Services (Verket för högskoleservice) that provides statistics on the number of applicants to Swedish HEIs.

Data on HEI annual *expenditure, funding sources, personnel, students and graduates* is collected regularly by the Swedish National Agency for Higher Education. Considering *publications*, a bibliometric database is developed by the Swedish Research Council (VR). Various efforts have been taken to allocate number of papers and citation to universities and subjects but work is not yet finalized. To obtain data on *patents* the Swedish Governmental Agency for Innovation Systems (VINNOVA) has access to PATSTAT, which can be used for time-series analysis. It is not obvious whether data can be allocated to the level of single universities. Individual universities may report on the number of granted patents in their annual report. Other source for patents is the KEINS database that contains statistics on patent applications in Sweden as well as in France, Italy and US reclassified by the name of applicant and inventor. There is currently no collection on overall and regular *technology transfer* data for all Swedish HEIs. In order to monitor technology transfer activities some of the large universities have collected data for specific years on university start-ups, patents and incomes from licensing. The Swedish National Agency for Higher Education conducts special studies on *employability* of graduates in Swedish municipalities disaggregated by universities.

Switzerland

Since the ’80s, there is a central statistical database on higher education institutions, based on regular data collection from the individual institutions. This source could provide data for *highest degree, students* and *research activity*. Descriptors can be easily provided from institutional websites.

Data on HEI *expenditure, funding sources, personnel, students* and *graduates* for ISCED 5A institutions is regularly collected by the Swiss federal statistical office directly from the involved institutions and published individually for each institution. For ISCED 5B institution, there is no systematic data collection on educational *expenditure* and *incomes*. Data on Web

of Science *publications* of Swiss institutions was produced and analysed until 2005-2006 by the Centre d'Etudes sur la science et la Technologie (CEST), being part of the Swiss Science and Technology Council. There is no systematic data on *patents* filed by Swiss universities. The Swiss federal office of statistics published a few data extracted from international databases, but disaggregation by institution is not provided. There is fairly good data on the *employment* situation of graduates from higher education institutions collected through a regular survey performed by the federal office of statistics one and five years after the diploma.

United Kingdom

The Higher Education Statistical Agency (HESA) is the agency responsible for most data collection for the UK higher education sector. HESA is a private limited company, which has formal agreements with government departments to provide the data they require, and it is funded by subscription from all of the universities and higher education colleges throughout the United Kingdom. All indicators could be obtained from this source (*region, highest degree, students, doctorates, research activity, specialization),* while for descriptors each institution should be contacted.

Data on *expenditure* and *funding sources* are available from HESA's Annual Finance Statistics Return (FSR) and need to be purchased. The staff record provides data in respect of the characteristics of members of all academic and non-academic *personnel* employed under a contract of employment by a HEI. The relevant CD-rom (Resources of Higher Education Institutions) has been purchased. The HESA Student Record collected data on *students*. Postdoctoral students are not included. The CD-rom "Students in Higher Education Institutions" also includes information on the qualifications obtained by students at reporting institutions. Data on *publications* by UK universities are available from Evidence Ltd. They need to be purchased. Data on income from intellectual property rights is available from HESA: Annual Finance Statistics Return (FSR). The HE-BCI (Higher Education Business and Community Interaction) survey, managed by the Higher Education Funding Council for England (HEFCE), reports a wide range of variables on UK universities' *technology transfer activities*. Data is public and freely available online. The HESA Destinations of Leavers from Higher Education (DLHE) target population contains data that could be used to elaborate *employability* indicators.

Summary overview

In each country there are main owners for data on higher education but for the extended set of data there are no official statistical sources for some indicators (as outputs) and other sources have to be analyzed. Table 36 displays the main sources of core set indicators, as well as some alternative sources available.

Table 36. Summary overview of owners of data by country

Country	Unit	Other Sources
Austria	Ministry of Research & Science, National Statistical Office	Each institution (data for *year of creation*)
Belgium	Observatoire de l`ensegnement superieur, Administratie Onderwijs en Vorming, Ministerium der Deutschprachigen Gemeinchaft, Belspo	Cref, Vlir, each HEI
Bulgaria	Ministry of Education & Science, National Statistical Office	BULSTAD register, DAKSI register, Ministry of Finance
Cyprus	Ministry of Education and Culture Statistical Office	Ministry of Trade, Industry and Tourism (patents), Each institution (publications)
Czech Republic	Institute for Information on Education Ministry of Education, Youth and Sport	Czech Statistical Office, R&D Council, Ministry of Labour and Social Affairs
Denmark	Statistic Denmark, Danish universities Association	Ministry of Sciences (data of publications and patents)
Estonia	Ministry of Education & Research	
Finland	Statistics Finland (KOTA, AMKOTA)	Each institution (*year, hospital, distance education, patens*)
France	Ministry of Higher Education & Research	BETA, CEREQ, OST
Germany	Federal Statistical Office	Centrum für Hochschulentwicklung
Greece	National Statistical Service Ministry of Education	Each institution (data for *year of creation*), General Secretariat for Research and Technology
Hungary	Ministry of Education and Culture Central Statistical Office (HCSO)	Each institution (data for *year of creation*)
Ireland	Higher Education Authority	Evidence Ltd.
Italy	Italian Ministry of Higher Education	Each institution (data for university hospital and year)
Latvia	Central Statistical Office of Latvia, Ministry of Education & Sciences	Centre for Science and Technology Studies (publications), Patent Office of the Republic of Latvia
Lithuania	Ministry of Education & Science National Statistical Office	Research & Study Monitoring and Analysis Centre, Registry Centre
Luxembourg	National Statistical Office, University of Luxembourg	ISI (publications), PATSTAT, EPO, USPTO (patents)
Malta	National Statistical Office, University of Malta	NCHE collects data from each institution, Ministry of Finances
Netherlands	Dutch national statistical office Ministry of Culture, Education & Science, Association of Research Universities	Central Funding of Institutions Agency, CWTS, Research Centre for Education and the Labour Market
Norway	Ministry of Education & Research Statistics Norway Norwegian Institute for Studies in Innovation, Research and Education	NIFU STEP (publications, R&D activities)
Poland	Central Statistical Office	The Information Processing Centre (outputs)
Portugal	Ministry of Science, Technology and Higher Education	
Romania	National Statistical Office	National Council for financing of higher education
Slovakia	Ministry of Education of the Slovak Republic (Data is administrated by the Institute of Information and Prognoses of Education), Statistical Office of the Slovak Republic.	Ministry of Finance of the Slovak Republic; Ministry of Labour, Social Affairs and Family of the Slovak Republic, Centre of Labour, Social Affairs, Family Centre of Scientific and Technical Information
Slovenia	Statistical Office of the Republic of Slovenia	Institute of Information Science (IZUM) (outputs)
Spain	National Institute of Statistics Vice-Chancellors Conference of the Universities,	Council of university coordination, ANECA, Spanish Universities Transference Office, WoS, EPO
Sweden	Swedish National Agency for Higher Education Statistics Sweden	Swedish Research Council (publications), Swedish Governmental Agency for Innovation Systems (patents)
Switzerland	Statistical Office	Centre d'Etudes sur la science et la Technologie (Publications)
United Kingdom	Higher Education Statistical Agency	Each institution, Evidence Ltd.

4.6 Procedures for data collection

The previous discussion allows developing some ideas and proposals concerning the regular data collection. Before this we will give a presentation of the procedures of data collection, integration and quality assurance applied in the EUMIDA feasibility study as this will also give some input to the discussion on future actions.

4.6.1 Procedures for data collection

The EUMIDA data collection was divided into two parts:

Data Collection 1: the core set of data for all HEIs, initiated on 26.01.2010.

Data Collection 2: the extended set of data for research active institutions, initiated on 30.04.2010.

The data collections were organized by use of templates developed in Microsoft Excel, in principle one variable and its metadata per sheet. The sheets were organized in a way that the information entered to identify the institutions was automatically repeated in each sheet to avoid confusion about the sequence of institutions. Accompanying the templates were guidelines with the data definitions, instructions for formatting and some additional standard definitions (NUTS codes, ISO country codes) for use in some variables.

As an alternative to templates, the use of an interface directly to the database through the web was discussed; giving country experts and statistical offices the opportunity to enter data directly into the database without the use of templates. This option was discarded as too risky in terms of lack of coherence and less control of data integration. In other circumstances, such as a permanent data collection done by experienced personnel, this option could be feasible.

Upon reception and quality checks of data (see below) data were integrated in a Microsoft Access database. This is a simple relational database, without the advantages of larger databases, but sufficiently powerful for the EUMIDA purpose. Besides, it has the advantage of easy integration with Excel as well as being available for many users through the Office software package. The database was constructed with tables very much resembling the sheets of the templates, facilitating easy import of data. Besides, tables were made for standardized values for uses in other tables and queries made for extraction of data. For inspection and analysis data were exported to Excel. In the database each institution was given its unique identifier (EUMIDA_ID), and this identifier was used across all tables to integrate the variables with the information identifying the institution (identifiers like name, country, region, year of foundation).

The tables in the database were constructed by use of a standard set of fields.

Table 37. Database field and content

Database field	Content
ID	Automatically defined ID for each posting in the table.
EUMIDA_ID	The specific ID for each institution in the format of XXNNN were XX is the ISO code for each country and NNN is the number of institutions for the same country running from 001 as long as necessary.
Variable	The field containing the specific data (in some cases more than one field, e.g. the various types of funding; total, core, third party etc).
Comments	A text field where comments to data in general or the specific institution can be inserted.
Reference year	The year for which the particular variable is valid.
Data flag	Specific information to explain the reason for missing data ('Not available', Not applicable') or problems with data ('Provisional value', 'Uncertain data' etc).
Data source	The original source of data.
Date of data collection	The date when data were collected for EUMIDA purposes.

Source: EUMIDA

Integration of data into the database was mostly done by 'copy – paste', while the initial data for the core data set were automatically imported. Automatic import of course saves work, while it reduces to some extent the quality control implicit in manual operations. For an experimental data collection like EUMIDA, manual integration gives a hands-on possibility for checking data for errors and consistency that is lost in automated systems. After most data were integrated, smaller revisions were done directly in the database as revised data came in.

4.6.2 Quality checks

EUMIDA applied a six-step quality check:

1. Data was collected from national statistical offices or by country experts familiar with the data production context of HEIs in their countries, and thus able to check their own data for errors and unlikely values. EUMIDA instructed national experts to carry out the first line of quality check.
2. Within the EUMIDA project group, each member was given the responsibility for a number of countries, including communication with the country experts on particular issues regarding the data as well as bringing issues to discussion in the whole group if deemed necessary. This ensured a coherent answer to questions raised on the interpretation of guidelines, choice of methodologies etc.
3. Data was collected and integrated into the database centrally. This made possible a check of coherence, identification of missing data, cleaning and completing data in a

coherent manner, integrating the data in the database by use of unique identifiers etc.

4. The integrated data was inspected by two quality managers, both renowned for their capacity in statistics and experience of analyzing data on higher education institutions. They inspected separately Data Collections 1 and 2 and focused their inspection on identifying outliers, skewed distributions between countries and other anomalies in the data (e.g. totals not corresponding to their constituent parts). Their reports resulted in questions to be answered by the country experts, resulting in further revisions and improvement of data.
5. The integrated country datasets were sent back to representatives for national statistical offices to check if data appeared correctly in the integrated dataset.
6. Through the analysis of data by the project team (as in this Report) further suspicious or problematic data was discovered and revised after consultations with country experts.

Being a feasibility study exploring the possibilities of building a coherent data set for European higher education institutions, we are reasonably sure that the overall quality through these steps has reached a level not seen in previous data collections, especially when taking into account the large number of countries and institutions in the data. This of course comes at the price of workload and time – as revisions in data have been made until the finalization of this Report. Future data collections have to be organized in ways taking greater advantage of automatic quality controls based on expected values or range of values.

4.6.3 Cleaning and completing data

To avoid unnecessary work on reformatting data, but foremost to ensure a coherent format of data for all countries, instructions were given on how to format data in the guidelines accompanying the data collection. Despite this, it turned out that in quite many cases data had other formats or there were missing data or metadata without any explanations, and data had to be further investigated and/or corrected before it were integrated in the database. The need for corrections can probably partly be explained by the fact that in some countries a large part of the data was collected by people that were not statisticians or used to work on data standardization; they were chosen also for other competencies. Most probably, quality will increase when data collection will become regular. However, in the first round of regular data collection one should expect a relatively large workload in cleaning and quality checks, as data on the individual level requires a high level of quality to be accepted also by the institutions themselves.

Some examples of problems and corrections, also useful as a memo for future data collections:

No data and no flag to indicate the reason for missing data: In many cases there were neither data nor and explanation to the missing data. In some cases it was clear from accompanying information that the statistical office intended to come back with further data at a later point in time, and had just omitted the variable or a number of institutions. In some cases data was later completed, while in other cases investigations had to be done to be able to insert more precise reasons for missing data.

Flags missing: In the guidelines instructions were given for how to use standardized flags to annotate characteristics of data, in line with normal statistical procedures, e.g. 'Not available', 'Not applicable' etc. in many cases these flags were not used, instead were given general comments in accompanying documents or emails which had to be transferred to metadata. In such cases the appropriate flag was inserted for all cases.

Inconsistencies between variables: Some EUMIDA variables are interdependent, in the sense that a certain value on one variable makes another variable meaningless. Typically it is the value 'Master degree' for highest degree awarded in combination with a positive number of doctorate degrees awarded. Technically, the flag 'Not applicable' should in such cases have been used for the variable *Number of doctorate degrees awarded*, indicating that this variable has no meaning for the institution in question. However, often this variable was left blank and the appropriate flag had to be inserted. More interesting is maybe the deviating cases were institutions not awarding a doctorate still have students at ISCED 6 level (doctorate). In quite many cases this is a real combination as the institutions may have teaching at this level, while the degree is delivered by another institution. This is thus an example of the need for thorough checking of each case.

Inserting comments and explanations: For all EUMIDA variables it was possible, as part of metadata, to give comments and explanations to the variables at the level of individual institutions. In some cases it turned out that such comments were inserted only for one of country's institutions, while it actually applied to all. In other cases such comments were pasted into the field for flags.

Standardizing formats: Before data collection great care was taken in developing a few standardized formats and drop-down lists of allowed values. Still, with many agencies and people involved, not all these rules were obeyed, and quite a few variables were reported using various formats. Typically it is reporting in thousands or millions where the exact data wase asked for, using various formats for date of data collection (dd.mm.yyyy, dd/mm/yy or name of month.yy, etc) or denomination of currency (EUR, euro, €).

Excel function of dragging cells: The particular function in Excel of dragging the handle of a cell to copy its value was sometimes needed for correcting data. As this function can copy the value and also fill in a series (N+1), it was sometimes obvious that the latter function had been used instead of the first, typically for dates of data collection increasing by one for each institution.

Ideally the workload for integration and quality checks should have been possible to estimate. However, since the procedures have involved many people, first in each country and second within the EUMIDA team, it was not possible to give an estimate. Below figures are presented on the estimated effort for collecting the data in each country.

4.6.4 Future data collection

In most countries there is a single organization owning most of the data in the core set and this should be considered as the organization in charge of core set of data collection, maintenance and delivery to EUROSTAT for the operational phase. We will refer to this unit in the following as the *national data collection unit*.

While for the core set it was possible to identify a single actor at national level – in most cases the National Statistical Authority, NSA – as owner of most data (except a number of descriptors), the situation for the extended dataset is more complex and differentiated by countries. Hence, after validation of this information, detailed scenarios for each country will have to be defined.

Some cases where the situation has to be clarified:

France: there is a multiplicity of owners of information in Higher Education and the source of data is related with the institutional dependence of each institution (different Ministries). This dependence must be considered for data collection.

Lithuania: there are different owners but data on students, degrees and research staff can be accessed on-line on MOSTA website (http://www.mosta.lt). The aggregated data is available for public consultation online from the Department of Statistics to the Government of the Republic of Lithuania and can be viewed either in pre-defined tables (http://www.stat.gov.lt/en/pages/view/?id=1594) or by accessing the database manually.

Malta: most data could be obtained from National Statistical Office but for the incomplete information it is recommended to collect data directly through each HEI. This is the better option due to the low number of institutions.

Netherlands: the data considered in the core set could be collected from several sources: the Dutch national statistical office: Statistics Netherlands; The Ministry of Education, Culture and Science or some if its agencies, in particular CFI, the Central Funding of Institutions Agency; IBG, the Information Management Group, an agency that handles the registration of (prospective) students and recognized degree programs, VSNU, the

Association of Universities in the Netherlands, HBO-raad, the Netherlands Association of Universities of Applies Sciences, The CWTS and NOWT, both being university-based research groups that publish information on the Dutch science system.

The following Table 38 provides an overview of these units. We notice that in most cases this is the same unit in charge of UOE data collection.

Table 38. Institution responsible for data collection in each country

Country	Unit	Remarks
Austria	Ministry of Research & Science, National Statistical Office	
Belgium	Observatoire de l`ensegnement superieur, Administratie Onderwijs en Vorming, Ministerium der Deutschprachigen Gemeinchaft, Belspo, Cref, Vlir	
Bulgaria	National Statistical Office Ministry of Education & Science	
Cyprus	Ministry of Education and Culture Statistical Office	
Czech Republic	Institute for Information on Education Czech Statistical Office	
Denmark	Statistics Denmark	
Estonia	Ministry of Education & Research Statistic Estonia	
Finland	Statistics Finland	
France	Ministry of Higher Education & Research	Owner is related to institutional dependence of each HEI
Germany	Federal Statistical Office	
Greece	National Statistical Service Ministry of Education	
Hungary	Ministry of Education an Culture Central Statistical Office (HCSO)	
Ireland	Higher Education Authority	
Italy	Italian Ministry of Higher Education	
Latvia	Central Statistical Office of Latvia	
Lithuania	Ministry of Education & Science	Data must be collected from different owners but MOSTA database provide some indicators
Luxembourg	National Statistical Office	
Malta	National Statistical Office	NCHE collects data from each institution
Netherlands	Dutch national statistical office Ministry of Culture, Education & Science	Different owners will be considered
Norway	Ministry of Education & Research Statistics Norway Norwegian Institute for Studies in Innovation, Research and Education	
Poland	Central Statistical Office	
Portugal	Ministry of Science, Technology and Higher Education	

Country	Unit	Remarks
Romania	National Statistical Office	
Slovakia	Institute of Information and Prognoses of Education and in some cases the Ministry of Education of the Slovak Republic	Statistical Office of the Slovak Republic will be acting as advisory body
Slovenia	Slovenian Office of Statistics	
Spain	National Institute of Statistics Council of university coordination	
Sweden	National Agency for Higher Education Statistics Sweden	
Switzerland	Statistical Office	
United Kingdom	Higher Education Statistical Agency	

Source: EUMIDA

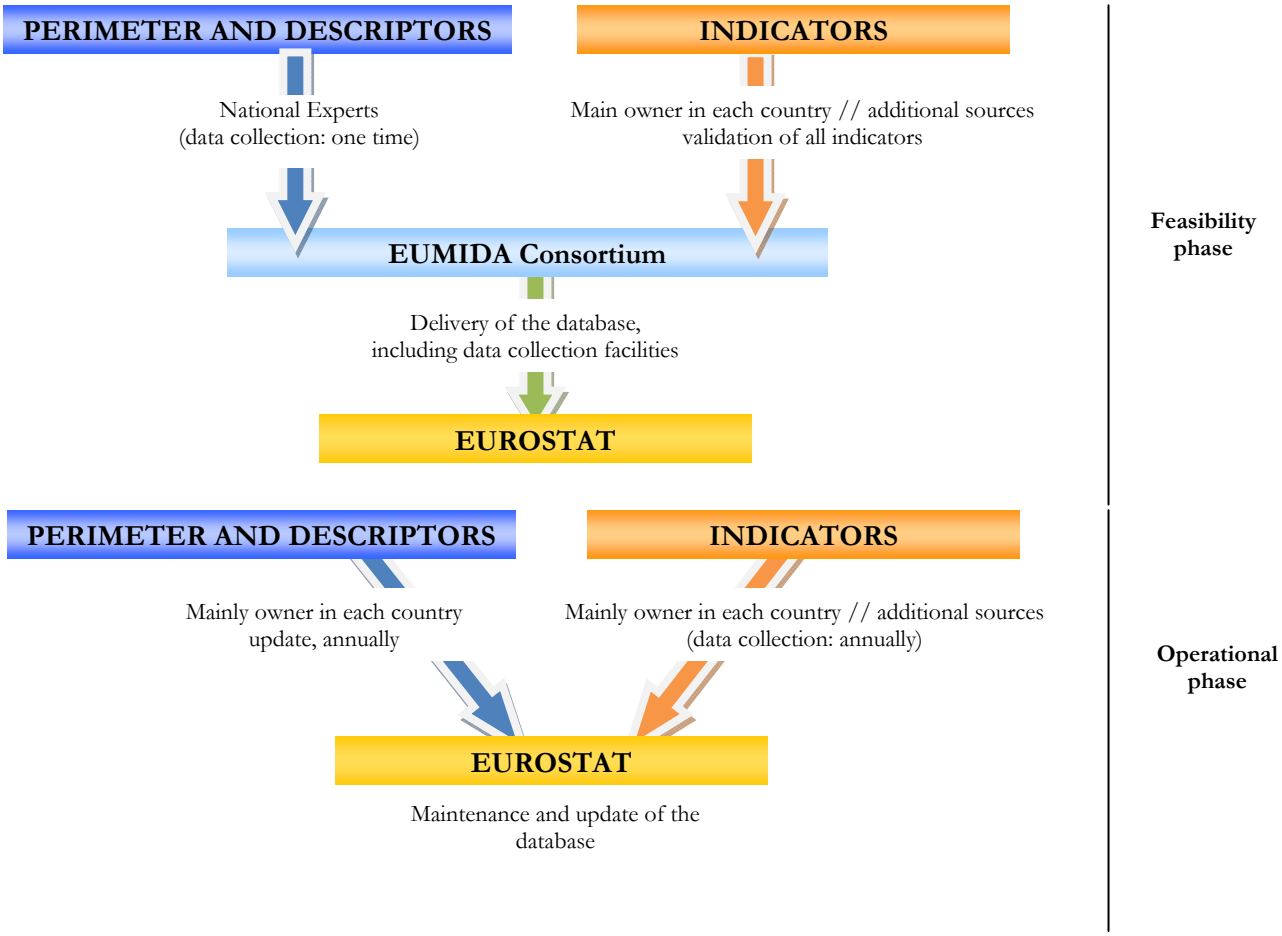
4.6.5 Data collection procedure

It is useful in this respect to distinguish between the definition of the perimeter, descriptors and data:

- The definition of the perimeter has in general been provided in the feasibility phase by EUMIDA national experts and validated by the national data collection unit. It needs to be updated annually by the national data collection unit, including coverage of demography (foundations, mergers, changes of status).
- Descriptors have been collected by EUMIDA national experts and validated by data collection units. This procedure will be important also in the future since *ad hoc* data collection is required in some cases, like changes in the status of the institutions. Descriptors will need to be checked annually by the national data collection unit and produced anew in case of changes of the perimeter.
- Quantitative data has been collected by EUMIDA national experts and validated by data collection unit. In the operational phase, they need to be produced annually by the national collection unit.

In the feasibility phase, all indicators collected were validated and verified by EUMIDA consortium with the aim of solving problems of availability and quality (see above). The resulting data set will be transmitted to EUROSTAT and constitute the basis for the annual data collection and updates by national contact unit. The following figure summarizes the procedure followed in the feasibility phase and proposed procedure for the future.

Figure 7. Procedure for data collection



Source: EUMIDA

4.6.6 Resources required and workload

The feasibility study for the 29 European countries shows wide differences in ownership of data, level of organization and structure of the sources of data for higher education in each country. Therefore it is not possible to calculate the time required to obtain data for the core set in all countries in general. It should be calculated for each country particularly.

To this step, the EUMIDA project developed a questionnaire to national statistical authorities to define more precisely the workload for annual data collection.

The collection of information proved more difficult than anticipated. A few NSAs declined to comment on the workload (e.g. Belgium, Germany, Italy, Lithuania, Luxembourg), while in other cases we felt the question was not properly addressed (Romania). However, in all other cases the national experts were able to collect quite detailed information. Table 39 summarizes the data collected. More detailed information with a breakdown by type of variable is available for several countries (not reported here).

Table 39. Estimated workload for a regular data collection

	Workload in man-days (estimate of additional workload to provide data for EUROSTAT under assumption that - if there are any - reasons for non-availability are overcome)					
Country	NATIONAL STATISTICAL INSTITUTION (responsible for UOE data collection; mainly National Statistical Institutes)			ALTERNATIVE SOURCE, IF IT IS BASIS FOR EUMIDA DATA COLLECTION (institution other than institution responsible for UOE data collection; e.g. Ministry of Research, Ministry of Finance, Rectors Conference, etc.)		
	Aggregate workload for data available and published	Aggregate workload for data available under request	Aggregate workload for data available after overcoming barrier	Aggregate workload for data available and published	Aggregate workload for data available under request	Aggregate workload for data available after overcoming barrier
Austria	2	8		1	2	
Belgium	n.a.					
Bulgaria				22	0	32
Cyprus	1	3	22	1	1	0
Czech Republic	6	9	13	1	4	4
Denmark	10					
Estonia	1	1	4			
Finland	18			8		
France						
Germany	n.a.					
Greece				3	77	
Hungary		3				
Ireland	3	5		3		
Italy	n.a.					
Latvia	1					
Lithuania	n.a.					
Luxembourg	n.a.					
Malta (*)					See footnote	See footnote
Netherlands	1			4	1	2
Norway	10			19		
Poland	60					
Portugal	45					
Romania (**)	n.a.					
Slovakia		44				
Slovenia	10					
Spain	37			57		
Sweden	15	6				
Switzerland	5					
United Kingdom	5			5		

(*) For Malta the reported data is in the two columns are 187 and 237, respectively, which must be clearly considered outliers.
(**) For Romania the reported data is 22 man-days for basic institutional descriptors and thousands of man-days for the rest of data, which must be clearly considered outlier.

The following remarks can be done:

- The estimated workload is not a function of the size of the country, but rather of the degree to which data is already available and tabulated on an individual basis;
- When data is available from NSAs, the range of estimated workload for most countries is between a few days and 2-3 weeks;
- There are three critical cases, namely Spain (37 days), Portugal (45 days) and Poland (60 days), which may require a careful consideration in terms of support from EUROSTAT;
- When data is not available from NSAs a more challenging situation emerges, that will require 2-3 months of additional workload in most cases (Spain, Norway, Greece, Bulgaria), with one outlier (Malta);
- In the cases where information were not provided by the NSAs, it is the opinion of national experts that the workload is in the order of weeks, rather than months.

In addition, at this stage, a few very preliminary indications can be given:

- Descriptors will be produced once and need just to be checked again each year; we foresee two major needs for revisions:
    - small scale changes for example with the foundation of new institutions or change of status of individual institutions (for example accreditation as universities). These are relatively unproblematic and less time consuming. Indicatively would expect a burden of a few hours for the smaller systems (less than 50 institutions) until a single day of work for the larger systems (200-300 institutions).
    - large reforms of the system with for example the creation of whole new sector of changes of status of whole group of institutions. Especially in large countries, this is likely to require an higher amount of work (indicatively a week of working time), but such reforms are relatively rare in most countries.
- Quantitative data need to be provided each year. These are currently six variables for each higher education institutions. In the countries where this data is available in some kind of databases and thus standard queries can be built the effort is likely to be very reasonable and not too much dependent on the number of institutions. An order of magnitude of one working day could be considered as reasonable. However, where data need to be compiled by hand from other sources, the burden could be much higher and thus cases need to be indentified carefully.
- A further workload of a few days work need to be provided for data control and merging and communication with EUROSTAT (including validation and questions).

The very preliminary conclusion is that the collection of data for the core set of variables requires a manageable amount of work at following conditions:

- That suitable standardised procedures and infrastructure are developed in the EUMIDA feasibility phase and that the units in charge of producing Data is directly involved in their preparation, so that these tools fit their requirements.
- That definitions and data collection procedures allow for use of existing national data without requiring much compilation and hand correction.
- Finally, that the required data for the six quantitative indicators is already available in a suitable electronic format and does not need to be collected institution by institution.

5. Comparability issues

5.1 Comparability and exploitation strategies

Once a methodology has been defined, including definitions and guidelines on how to measure the different variables, and data have been collected, as in the EUMIDA project, a central question concerns the level of comparability of the data. Namely, for all exploitation purposes of institutional-level higher education data – including scholarly analyses of higher education systems, political decisions at the national and European level, evaluation of individual institutions – it becomes central to check to which extent "numbers tell the truth" concerning the specific questions to which they are supposed to answer.

Concerning comparability, EUMIDA brings additional levels of complexity with respect to aggregated higher education and R&D statistics at the level of whole countries. First, the disaggregation at the institutional level means that measurement problems will be more severe, since the statistical units are now smaller, and institutional specificities might have larger impact on the data (a well-known phenomenon in statistics). Second, thanks to this disaggregation, data at the institutional level are likely to be used by a wider audience and thus comparability issues might become more sensible (e.g. if rectors use this database to compare their own universities with other and draw managerial conclusions). Third, the EUMIDA dataset has been purposefully developed to answer to a broad range of policy questions and thus it is likely to be used not only for purely descriptive comparisons – like comparing number of students across institutions -, but for more in-depth analyses, like measuring efficiency of institutions, comparing their research output, etc. As we shall discuss later, this kind of analyses raises comparability issues going well beyond the technical quality of the data themselves.

Also, many comparability problems are not of technical nature – meaning that these could be addressed through improvements in definitions and data collection – but of ontological nature, arising when we try to compare objects which don't have the same structure or abstract properties which cannot be unambiguously measured, e.g. because their measure closely depends on underlying actor's beliefs and goals.

This does not mean that one should distrust the EUMIDA dataset or avoid far-reaching analyses by using it; at the contrary, we believe that it constitutes a powerful instrument both to better analyze higher education systems, especially if coupled with other data sources like bibliometrics databases, as well as to support political decisions. However, its users should be aware of some basic attitudes that need to be taken when exploiting this data:

- First, the production of each indicator is closely dependent on underlying representations and beliefs, for example on what universities are, on their mission, on what is research and educational quality, etc.; care should be taken to explicit these choices and to pick

the right indicators corresponding to them. Without forgetting that the EUMIDA dataset itself has been based on a number of conceptual assumptions explained in Chapter 2. We thus discuss the epistemological nature of indicators and its implications for the EUMIDA dataset in Section 1 of this Chapter.

- Second, comparability is always a relative concept depending on choices on which characteristics are important, on the system's perimeter, on what we want to compare; thus, a discussion of comparability is always relative to the specific use of data; in some context and for some purposes some data might be sufficiently comparable – e.g. if comparability problems do not fundamentally change the rank among different institutions -, while this might not be the case in other contexts. We provide a more in-depth discussion of types of comparability problems related to higher education in Section 2 of this Chapter.
- Third, no database is free from mistakes and technical problems, be it for problems in the data collection itself, be it because of data which are not fully reliable; in practical terms, establishing a statistical database means also tolerating some level of diversity in the data, the alternative being having no database at all. However, some of these problems might have impact on further data analysis; this means that, first, consistency check with other kind of information is well-advised and, second, that metadata on comparability problems and statistical differences needs to be carefully checked each time. Section 3 provides some insights on these more technical problems and on differences in definitions in the EUMIDA dataset.
- Fourth, statistical data is measures deprived of their context and this is their value as a simplification of the reality; however, context needs to be reinserted in the analysis at a later stage in order to check if quantitative comparisons provide meaningful results also concerning the reality itself. This is a central issue in higher education, given the wide diversity both of national systems and of institutions themselves (enough to think to subject mix). Thus, in Section 4, we discuss some of the main contextual differences that need to be taken into account in using the EUMIDA dataset.
- Finally, in most cases indicators do not provide definitive and objective answer to policy or research questions, but only valuable evidence which needs then to be validated and discussed; indicators are not answers, but instruments to nurture the scholarly and political debate and one should refrain from using them to decide which position in that debate is the correct one. We finally discuss this role of indicators in the context of higher education in Section 5 of this Chapter.

5.2 What S&T indicators are: an introduction and some applications to higher education

Even if EUMIDA has developed basically a statistical database mostly containing simple descriptors – like numbers of students, it is clear that the main purpose of developing the

database was to provide some quantitative information to answer to far-reaching questions like: *What is the level of diversity of European higher education? Do we see convergence or path dependency in the level of diversity? Which are the most efficient institutions? Which institutions are active in research?* etc.

This is exactly the role of numbers qualified as *S&T indicators* (Moed *et al.* 2004; Barré 2004): they are constructs that allow the quantitative measurement of specific features of reality which cannot be directly observed. This is achieved through the combination of different data and the use of suitable proxies, which are supposed to be related to the observed feature. For example, the number of citations of scientific publications is taken as a proxy for the outcome of research activities, because scientific studies have extensively shown that (1) recognition by other scientists is a reasonable measure of research quality, and (2) this is expressed by scientists citing these works in their own papers.

Thus, indicators are based on some kind of conceptual modelling, which has to be rooted in science and technology studies, but also to some extent in the *definition and normative understanding of the underlying reality* and basic concepts. In our example, if one would assume that research quality is strongly related to the application potential of its outcome and resulting benefits for society, the bare use of academic recognition to measure it would have to be questioned and different indicators might be more appropriate. From this example it is obvious that in some sense indicators are socially constructed and their appropriateness is related to the specific perspective of their users and their underlying objectives and questions. The fact that they appear objective to most users and commentators depend on the accumulation of practice and consensus building among actors in society, coupled with increasing sophistication in technical treatment (Desrosières 1993). Thus, there are no indicators without a clear definition of the underlying questions (Barré 2004; Godin 2005; Lepori, Barré and Filliatreau 2008).

Moreover, indicators are explicitly meant to condensate information. A comprehensive assessment is replaced by less precise information which demands less effort and resources and which is easier to interpret because of its quantitative nature. For example, the production of an impact indicator requires much less effort than a complex peer-reviewed assessment of a scientific field. But the price of this is a loss of detail and depth. A thorough analysis would be more suitable to avoid misinterpretations – but at the price of a much higher effort. Therefore, a central criterion for a good indicator is its ability to provide a reasonably precise picture of the reality with an acceptable effort.

A closer look reveals however wide differences in the epistemological status from indicator to indicator, as well as considerable fuzziness of where to draw the distinction between data and indicators. Thus, at least in their origin, all types of data – including the most basic ones, like physical measures – are theory-based. But what matters is the purpose of their use as simple descriptors or as proxies for underlying phenomena.

It is useful to make some distinction between the following broad classes:

a) ***Descriptors*** limit themselves to describing some aspects of reality without adding any further interpretation. An example from EUMIDA is providing descriptive statistics on the number of students.

Most data contained in the EUMIDA dataset, used at their face value, are simple descriptors. They are subject to technical problems of measurement (see Section 3 in this Chapter), as well as to differences related to the context, but raise usually rather limited interpretation issues. Accordingly, also their explanatory value is more limited: it is useful contextual information to produce some descriptive statistics concerning the number of higher education institutions, the students, costs, etc., but the real value of this information will emerge when interpretations are added on their broader meaning.

b) ***Markers*** are used as proxies for phenomena that cannot be directly measured. For example, patents are considered as a typical marker for technological outputs. However, a marker does not attempt to deepen the understanding of the relationship between these two layers, e.g. with some sort of quantitative relationship.

An example of marker in the EUMIDA dataset is the *research-active variable*; it marks if an institution possessed institutionalized research activities. As discussed in Chapters 6 and 7, this marker is useful in that it discriminates between institutions that by and large don't have research and the rest of research active; however, it does not allow going deeper on the volume, quality or impact of research activities.

Markers already depend on conceptual assumptions on the observed phenomena; the criteria used for research-active institutions (see Chapter 2 of this Report) yield a much larger set of institution than, for example, adopting the doctorate as a criterion, even if the results produced by these two markers are not unrelated (see Table 40). In particular, there is a very small group of institutions that, although accredited with the authority to grand doctoral degrees in fact do not have PhD activities, or do not fit with the other criteria used to define research activity. At the other hand, there is a large group of research active institutions that however do not grant the PhD.

Table 40. Research active and doctorate-awarding institutions. Data from the subset (n= 2410) of units with full information.

	Doctorate awarding	Non doctorate awarding
Research-active	850	555
Non-research active	42	973

Source: EUMIDA

It should be clear that there is no "objective" choice in this respect; the choice made in EUMIDA was motivated by the wish of providing a broader view of the research mission of HEIs beyond those doctorate-awarding, assuming that also research in other types of institutions is relevant at the system level (e.g. in Universities of Applied Sciences; Kyvik and Lepori 2010).

c) Finally, ***indicators*** are constructs that explicitly build the connection between quantified information and non-observable properties. For example, citation impact indicators (normalized per field) are used as quantitative measures of scientific quality.

Indicators forcefully require the creation of some kind of benchmark or scale against which to measure the score of individual institutions. In turn, this will be highly dependent on conceptual choices, but also on the perimeter of the considered population, on the chosen benchmark approach and, finally, on contextual variables which might strongly impact on the observed relationship. Hence, to be trusted, indicators require extensive validation and fine-tuning by comparing their results with reality, as it has been done for example for impact measures through citations by correlating these scores with peers judgments.

A classic example in higher education statistics is the use of the ratio between doctoral degrees and undergraduate students as a measure of research intensity; this indicator is adopted in the US Carnegie classification to identify *research-intensive HEIs* (those with more than 1 PhD degree per 100 undergraduate students). It depends on the definition of research HEI as doctorate-awarding institution. This definition is not forcefully adequate to the European context. What is more important, this indicator is dependent on the subject mix – with natural sciences having systematically more PhD students per undergraduate student than social sciences – and on the country, with HEIs in countries like Switzerland or UK having systematically more PhD students than those in Italy or Spain (see Chapters 6 and 7) (Bonaccorsi 2009).

Of course, it might be that these differences reflect real differences in the level of research-intensity across countries and subject domains, but one might wonder if to some extent these relate to other factors, like different form of organization of research or a different role of the doctorate in the society overall. Concerning indicators, careful calibration and control of contextual effects thus becomes of prime importance.

5.3 Data and indicators comparability: a discussion of the concept and some implications

A second important issue concerns the meaning of comparability itself. At first face this would seem quite simple, as data can be considered as comparable if and only if for all couples of HEIs which have in reality the same characteristics the corresponding measure is the same (within measurement errors).

However, this positivistic concept of comparability drawn from experimental sciences can hardly be applied to S&T indicators, which in most cases deal with properties which are too abstract to be precisely and objectively measured, as discussed in the previous Section. And one might doubt if any of the indicators used in science policy would satisfy this criterion.

A more usable concept of comparability relates it to the purpose for which the indicators are used: thus, indicators are comparable if they yield results corresponding to the reality in a given conceptual framework and classification scheme.

A simple example on university rankings will clarify this discussion. These rankings are heavily criticized because of their underlying assumptions – e.g. the focus on international academic visibility- as well as because the ranks of individual institutions heavily depend on the chosen ranking and vary from year to year. In this perspective rankings are certainly not comparable and do not provide a reliable ordering of HEIs worldwide concerning their research quality.

However, under some assumptions, rankings do provide valuable results:

- If it is accepted that international scholarly visibility is the main component of research quality.
- If the use of rankings is limited to the top-200 HEIs in the lists.
- If they are used to identify groups of institutions rather than the relative position of individual institutions.

Under these conditions, experts would agree that most rankings identify HEIs with high international visibility, as well as the core of the internationally leading research universities worldwide and, in fact, in this respects different rankings yield reasonably similar results (even if results for individual institutions and years might diverge). It is generally recognized that results of different rankings converge if applied to the upper tail of the distribution, while have little correlation below a certain rank.

A further issue about comparability relates to the purpose of the comparison and to its implications. If comparisons are done for intellectual and scholarly reasons – e.g. to analyze the European university system and draws some general conclusions – requests on comparability will be much less stringent than if these are done for evaluative purposes – e.g. to evaluation research quality of a HEI – or to take political decisions – e.g. on the rules for the whole system – or to distribute funding; thus, all countries using HEI data for the distribution of resources have introduced well-developed data collection and control procedures, exactly because of the practical implications of measurement errors. The timeframe is also important: if measures can be revised and improved later their comparability will be less critical than if indicators provide direct and irreversible effects, e.g. closing down a university because of evidence of insufficient performance (as measured by some kind of indicator).

This *relative notion of comparability* has important implications for the use of EUMIDA data. As documented in Chapter 4 of this Report, EUMIDA made a large effort in constructing the best possible database of HEIs at the European level building on existing data from national statistical systems that have been developed for different purposes and with different underlying conceptions of higher education (see the following Section). Moreover, this database is experimental, as it has constructed in a pilot project, and thus could not include feedback from extensive use of data on any comparability problems.

Based on the analyses presented in this Report and an extensive validation with expert knowledge, we consider that EUMIDA data provide a fairly correct picture of the overall system and its institutional diversity, e.g. by types of institutions, by subject, by type of

activity and we expect that the broad categorizations provided – e.g. research active vs. non-research active – are by and large correct. Users should however be much more careful in comparing directly individual institutions, beyond the comparison between the simplest descriptors and, in all cases, we consider that this data should not be used directly for policy purposes without a through validation and discussion with experts (see further in the last Section of this Chapter).

5.4 Technical issues and their impact on comparability

Besides general epistemological and conceptual questions, problems of comparability are related to more specific technical issues concerning data quality and availability, related also to characteristics of national statistical systems. Of course, many technical questions are related to differences in underlying conceptions or in the national context; nevertheless, it is important to deal specifically with these more specific aspects, as they emerged in EUMIDA, in order to design robust exploitation strategies.

A complete list of all comparability problems emerging from the data collection has been provided as an annex to the database in a specific file, where deviant cases, explanations to data and missing cases are explained by national experts; accordingly, flags have been inserted in the dataset itself. For all kind of use of this data, a careful analysis of these metadata is recommended.

While Chapter 4 of this Report provides a more in-depth analysis of data availability by country, we focus here on a number of general comparability problems in the EUMIDA database, as well as on a few examples of more specific (country-level) issues.

5.4.1 General comparability issues

Table 41(a) provides an overview of some general comparability problems that emerged in the EUMIDA data collection, of the extent to which they impact on the analysis and some possible solutions we suggest at the level of data exploitation. Hence, the strategy suggested here is to pick among the exploitation strategies – e.g. the construction of indicators, the benchmarks used etc. - those which are less influenced by data problems. Of course, in the long term, some of these problems could be addressed by improved data collection procedures and some standardization of definition and rules (e.g. concerning accounting).

Table 41(a). Some relevant general comparability problems and their possible solutions

Problem	Description and data affected	Strategies suggested
Differences in the national perimeter	Coverage of ISCED 5 and 6 is different according to the countries and this affects the total number of institutions in the system. National averages per institution can be strongly affected since differences in coverage mostly pertain to smaller institutions. Also distributions of institutions by class can be strongly affected (e.g. breakdown between public and private institutions).	Use medians instead of average for cross-country comparisons or weight comparisons by number of students or staff. Set thresholds (e.g. exclude all institutions with small number of students). Limit comparisons to categories of institutions where coverage is better (e.g. public-sector).
Missing data for some variables	Even if institutions are included in the perimeter, data for some variables can be missing for some of them (see details in Chapter 4). Thus, even if overall coverage of countries is broadly comparable, this might not be the case for specific variables (e.g. expenditure data) and this is likely to affect comparisons.	Check detailed data availability for each variable. Adopt statistical correction techniques for differences in coverage of individual variables. Limit comparisons to categories of institutions where coverage is more uniform for the specific variables used.
Differences in price levels between countries	Since the level of prices and wages is very different between countries, the same face value of revenues and expenditure have different meaning in each country in terms of the ability of institutions to purchase resources (e.g. to hire staff). This does not affect only countries outside the Euro zone, as price levels are also very different in the Euro zone; this affects especially Nordic countries and CEEC.	Convert all monetary values for cross-country comparison by using Purchasing Power Parities, *including for countries using the euro*.
Differences in wage levels in the higher education sector	Despite the use of PPPs, differences between countries can arise because of different wage levels for academic personnel, which are not reflected in PPPs calculations (as based on a general basket of services).	Use alongside comparisons of financial data, comparisons based on staff data, which are usually more comparable and reliable.
Capital expenditure	Data on capital expenditure are not available for many countries (see Chapter 4) and, when available, they are not comparable because of different accounting systems in each country. Even comparability between institutions in the same country might be highly problematic.	It is suggested not to use the data in EUMIDA for cross-country comparisons, but to use only current expenditure data. Use of flat rates by subject domain could be a strategy to take into account different levels of capital expenditure by domain.
Staff data in headcounts rather than in FTEs	Quite some countries in EUMIDA were able to provide staff data in headcounts. This might lead to strong distortion in comparisons as there is evidence that the share of part-time staff differs between countries and types of institutions (e.g. being much higher in some colleges). Other countries use pro-rata (e.g. full-time staff + 50% of part-time in Poland).	No real solution except to clearly label the data. Comparisons should be limited to similar types of institutions to reduce the impact of these differences.
Status of PhD students	In some countries most PhD students are under regular contract with the university and thus included in staff numbers, while in other countries they are considered as students and either not paid overall or funded through national grants (and thus not included in staff). This affects all analyses based on staff numbers, like student to staff ratios, productivity indexes etc.	This a major comparability problem not having an easy solution without changes in data collection at national level.

Source: EUMIDA

We focus here on the mandatory data collection (both Data Collection 1 and 2), while the reader should refer to Chapter 3 for further information on methodological problems concerning research output data.

5.4.2 Country-level comparability issues

The following Table 41(b) provides some information on a few specific comparability problems related to individual countries. This is not to imply that work in EUMIDA has been performed better in some countries than others or statistical systems are of better quality;

as it shall be clear, most of these problems are related to specific institutional settings at national level. Also, this list is only exemplary and the users of EUMIDA data need to carefully consult metadata in order to get complete information by country and data type.

Table 41 (b) offers a reasonably complete "guided tour" across the countries in search of problems of comparability. We emphasize this is an important contribution, since one of the common themes in the debate around the European landscape in higher education is the "impossibility" of comparison. Comparisons are always far from perfect, but the only way to address the problem is disentangling all possible sources of non-comparability.

We believe one thing is the recognition of heterogeneity in the landscape, another is the claim that any quantitative comparison is, in principle, flawed and ultimately misleading. Quite on the contrary, it is the progressive build up of a common statistical basis that will permit to identify and delimit all definitional, qualitative or measurement issues that must be taken into account in comparing institutions across countries.

Table 41(b). Some comparability problems specific to individual countries

Problem	Country	Description and data affected	Strategies suggested
Coverage of Health and welfare	Netherlands	Health & Welfare education in the Netherlands has been integrated in the Hospitals, and thus data from this country do not cover students at ISCED 6 level and staff for this sector in universities (while ISCED 5 students are included). This affects all comparisons between Dutch and other countries institutions in terms of size, performance, etc.	Either compare by excluding this domain or use estimates based on flat rates (e.g. number of staff per student).
Diverging classification of students	Spain	ISCED 5 students are classified by five domains only and thus students of the other domains are included in these domains (e.g. education in social sciences). This will affect all comparisons concerning subject mix between Spanish institutions and other countries.	Groups subjects in other countries or use estimates to correct for the missing domains.
Diverging classification of students	Netherlands	Slight differences between the Netherland's classification and the UOE classification (e.g. informatics included in engineering).	Check case by case if it impacts on the results of the analysis.
Different period of reference for financial data	Ireland	Financial data refer to the period 1.10-30.09. If the evolution if institutions is smooth, this is unlikely to affect comparisons.	Check if any changes (e.g. mergers) occurred and, in case, correct individual data for purpose of comparison.
Inclusion of academy of sciences	Bulgaria	Bulgaria included the National Academy of Science in the perimeter because it is the largest provider of doctoral education in the whole country. However, since it has no undergraduate educational activities, this will bias all kind of comparisons like of students/staff ratios, share of R&D activities etc.	At the institutional level, the Academy of Science should be clearly marked as an outlier and should be excluded from comparisons dealing with educational activities.
Missing breakdown of income	Czech Republic	Breakdown of income by core and third party is not possible for all revenues, since universities research plans do not fit into this categorization. This affects comparisons on the composition of revenues.	Czech Republic should be clearly flagged in comparisons based on categories of revenues.
Classification of staff	Finland	Breakdown by FOS is based on headcount and thus is not comparable with aggregated data (in FTE).	For purposes of data analysis correction is suggested to use the relationship FTE/headcounts for each institution as correction of the data.
Classification of staff	Norway	Part of the staff cannot be attributed to FOS categories and thus sum of the categories does not add to the total.	Distribute this staff pro-rata on FOS categories when required for analysis.
Different year of reference	Spain	Latest available data refers to 2006	This should be addressed in the regular data collection; check case by case if it influences comparisons and flag results.

Source: EUMIDA

5.4.3 Comparability of fields of education and fields of science

As noted in Table 41(b) one of the areas where comparability across countries is not perfect is the classification of students into fields of education. How severe is this problem? Luckily enough, the classification of fields of education is one the areas where the international

statistical systems have been working more intensely for years, so that classifications are quite robust and updated regularly. Quite another problem, which must be mentioned here, is the relation between fields of education, as classified in the UOE Manual, and fields of science, as classified in the Frascati Manual.

We noticed that there is no single scheme of classification for all EUMIDA variables, since these represent different realities and have to be classified using different principles; hence, the EUMIDA approach was to keep the original classification schemes and to build afterwards correspondence tables where this is required.

Broadly speaking, EUMIDA has adopted the following classification schemes:

- the classification by fields of education of the UOE manual (UOE manual 4.2), which is based on the contents of the educational programmes; it thus used to classify programs, students and degrees (based on the program where they are enrolled);
- the fields of science and technology (FOS) classification of the Frascati manual (Frascati manual 3.6.2), which is based on fields of science where R&D activities are performed; EUMIDA has extensively checked its usability for classification of academic personnel (as alternative to FOE).

For students and degrees, data should be divided by fields, adopting the first level (broad fields) of the fields of education classification of the UOE manual (UOE 4.2), since this is classification is widely for data in educational statistics. The classification distinguishes following fields at levels 1 and 2.[21]

[21] For full details on the classification please refer to the specific Fields of Education and Training Manual (Eurostat 1999; http://circa.europa.eu/Public/irc/dsis/edtcs/library?l=/public/measuring_lifelong/classifications/isced97_fields&vm=detailed&sb=Title).

Table 42(a). Fields of education classification

ISC 1	Education	Teacher training (ISC 141) Education science (ISC 142).
ISC 2	Humanities and Arts	Arts (ISC 21) Humanities (ISC 22).
ISC 3	Social sciences, business and law	Social and behavioural science (ISC 31). Journalism and information (ISC 32). Business and administration (ISC 34). Law (ISC 38).
ISC 4	Science	Life sciences (ISC 42). Physical sciences (ISC 44). Mathematics and statistics (ISC 46). Computing (ISC 48).
ISC 5	Engineering, manufacturing and construction	Engineering and engineering trades (ISC 52). Manufacturing and processing (ISC 54). Architecture and building (ISC 58).
ISC 6	Agriculture	Agriculture, forestry and fishery (ISC 62). Veterinary (ISC 64).
ISC 7	Health and welfare	Health (ISC 72). Social services (ISC 76).
ISC 8	Services	Personal services (ISC 81). Transport services (ISC 84). Environmental protection (ISC 85). Security services (ISC 86).

The classification by fields of science and technology (FOS), on the other hand, has been introduced in the Frascati manual in the '60 and slightly revised since then (Frascati manual, 3.6.2). The last revision of the FOS classification was conducted by OECD in 2006. Classification should be based on the scientific fields in which most of the activities of an institutional unit (for example a department or a research centres) are undertaken. The Frascati manual recommends to perform this classification at the level of subunits where activities are sufficiently homogeneous.

FOS is adopted to classify research expenditure and staff in R&D statistics and hence it has been considered as the scheme for classifying academic personnel. Table 42(b) provides an overview of the FOS classification.

Table 42(b). Fields of science classification

FOS 1	NATURAL SCIENCES	1.1 Mathematics 1.2 Computer and information sciences 1.3 Physical sciences 1.4 Chemical sciences 1.5 Earth and related environmental sciences 1.6 Biological sciences 1.7 Other natural sciences
FOS 2	ENGINEERING AND TECHNOLOGY	2.1 Civil engineering 2.2 Electrical engineering, electronic engineering, information engineering 2.3 Mechanical engineering 2.4 Chemical engineering 2.5 Materials engineering 2.6 Medical engineering 2.7 Environmental engineering 2.8 Environmental biotechnology 2.9 Industrial Biotechnology 2.10 Nano-technology 2.11 Other engineering and technologies
FOS 3	MEDICAL SCIENCES	3.1 Basic medicine 3.2 Clinical medicine 3.3 Health sciences 3.4 Health biotechnology 3.5 Other medical sciences
FOS 4	AGRICULTURAL SCIENCES	4.1 Agriculture, forestry, and fisheries 4.2 Animal and dairy science 4.3 Veterinary science 4.4 Agricultural biotechnology 4.5 Other agricultural sciences
FOS 5	SOCIAL SCIENCES	5.1 Psychology 5.2 Economics and business 5.3 Educational sciences 5.3 Sociology 5.5 Law 5.6 Political Science 5.7 Social and economic geography 5.8 Media and communications 5.7 Other social sciences
FOS 6	HUMANITIES	6.1 History and archaeology 6.2 Languages and literature 6.3 Philosophy, ethics and religion 6.4 Art (arts, history of arts, performing arts, music) 6.5 Other humanities

FOS and fields of education classification are based on different principles (research fields vs. content of educational programs) and thus classify different objects. Nevertheless, at the

1st level of classification the two classifications are broadly compatible, with some rather minor differences. The following table provides an overview.

Table 43. Correspondence between Fields of education and Fields of science

Fields of Education	Fields of Science
0. General programs	-
1. Education	5.3 Educational sciences
2. Humanities and Arts	6. Humanities
3. Social Sciences, Business and Law	5. Social sciences without 5.3 Educational sciences
4. Sciences	1. Natural sciences
5. Engineering, Manufacturing and Construction*	2. Engineering and technology
6. Agriculture	4. Agricultural sciences
7. Health and welfare	3. Medical sciences
8. Services	-

* Includes urban planning which in the FOS classification is included in 5. Social sciences

Since all countries follow these definitions quite strictly, the problems in comparability outlined in the previous Section can be considered the only ones relevant.

5.5 Reintroducing the context in higher education statistics

Some of the comparability problems are related to data availability and limitations of existing statistical systems, as when some kind of disaggregation of data is not available. However, other comparability problems come from structural differences between national systems or characteristics of individual HEIs, so that they can hardly be overcome by improved statistical practices. In these cases the same number, collected with the same statistical definition, may tell a different story according to the national context or the type of HEI. These cases call our attention on the need to contextualize data and available indicators in their institutional and national setting, on the one hand; and to take into account these specificities in the statistic and econometric analysis on the other hand.

For analytical purposes, it is useful to distinguish between two main sources of comparability problems, namely, differences in the *organization and governance structure of national higher education systems* and *heterogeneity of the individual HEIs*. We provide below a discussion of these issues with examples derived from the EUMIDA experience.

5.5.1 Institutional context

An important source of heterogeneity between universities comes from the ways institutional systems are conceived across European countries. There is a large body of literature on the system-level governance of higher education systems and on its changes during the last two to three decades (Clark 1983; Neave and van Vught 1994; Amaral *et al.* 2002).
We notice that, at least in federal countries, the institutional context can be heterogeneous even at national level as the Swiss case demonstrates (Lepori *et al.* 2007). With the increasing role of the regions in the European Research Area, the issue of the impact of regional differences – concerning levels of economic development, industrial structure, and support measures – is also increasingly important (Larédo 2003).
One should be aware that heterogeneity of institutional context is a prime feature of European higher education and a major difference with the United States and require particular care in cross-country comparisons. It is thus strongly suggested that quantitative analysis using the EUMIDA dataset is enriched by more qualitative information on the characteristics of the national systems, as available from the higher education literature or information systems like EURIDYCE.

Dual systems

Many European countries have a system in which the Higher education system includes universities but also non-PhD-awarding institutions, like Fachhochschulen in Germany, Hogescholen in The Netherlands and Universities of Applied Sciences in Switzerland (Huisman and Kaiser 2001; Kyvik 2004). In most cases, these HEIs differ clearly from universities concerning their organization, education and research output. In other countries (e.g. Norway), we see a transition of such institutions as they are obtaining the right to award PhDs within a limited set of scientific fields. The appropriate strategy here is to carry out all analyses separately and to specify models differently; since sector classification is not included in EUMIDA, as it is not a statistical information, care should be taken to retrieve this information from national descriptions when the analysis is performed.
However, the interactions between the two sectors should be considered, especially concerning education, since the existence of these "second type" institutions tends to somewhat lower the number of undergraduate students in universities. Moreover, in countries like Switzerland, Norway and Finland, these institutions are relevant players in "third mission" activities towards the private economy. When interpreting data on institutions in one sector, care should be taken to consider the impact of the existence (or non-existence) of other sectors.

Public Research Organizations

In some countries, a large part of scientific research is carried out not only by universities, but also by large Public Research Organizations (PROs), such as CNRS in France, Max Planck

in Germany, CNR in Italy, and CSIC in Spain. Thus, any analysis of HEIs should take carefully into account the relative role of universities and PROs in the production of research output, as well as the interaction between these actors in terms of joint production of research output.

This issue is particularly relevant for countries, like France and to some extent, Portugal, where mixed units are widespread and, hence, it is difficult to identify clearly the perimeter of the HEI. Thus in France joint laboratories between CNRS and universities include people funded by both institutions and research contracts might be managed through both sides; also, institutional affiliation of publication and patents is handled case by case and thus it is quite difficult to separate inputs and outputs between university and CNRS. There is no easy solution to this issue since it reflects an underlying organization of the research system. A solution that could be investigated is to consider CNRS as a funding agency and to include all joint laboratories in the university perimeter. However, this would require data collection from both sources.

Funding patterns

Heterogeneity is also introduced by differences in funding, since national funding systems are quite different across Europe concerning the level of funds and their composition. Thus, comparing levels of spending or the composition of funding (e.g. third-party vs core funding) across institutions and countries requires taking into account differences in the national level and composition of funding. The usual approach in this respect is to compare institutional variations with national variations; thus, from earlier work it known that the share of tuition fees in total revenues is essentially nationally determined (with the exception of the UK); whereas the share of grants and contracts shows much greater variations between individual institutions (Lepori *et al.* 2007). Adopting different benchmarks – for example normalizing the share of third-party funds against the national average – might be useful in this context.

5.5.2 Heterogeneity of individual HEIs

Another fundamental issue is the level of heterogeneity at the institutional level, as the result of differences in the institutional systems, as well as the outcome of historical developments and strategic decisions. EUMIDA provides some help in this context since the dataset includes a set of basic descriptors that are not usually included in statistical systems, but are highly relevant to interpret the results. These provide information on the legal status of the institution, age, presence of a university hospital, and distance education.

However, without in-depth knowledge of individual cases, disentangling these effects proves to be quite difficult.

Subject mix

Higher education institutions may have very different profiles in terms of faculties and schools they are composed of, and hence of educational activity and research output. From the teaching point of view, strong variations in cost per student are likely to occur, due to differences in capital intensity (laboratories), length of curricula, type of training (theoretical, applied, practical experience) and so on. As it has been shown in earlier work from the AQUAMETH project, differences in subject composition are in fact the main explaining factor of differences in staffing and costs per students (Lepori *et al.* 2010). Moreover, some of the most largely used indicators of research output are strongly dependent on subject domains, as the coverage of international publication databases is much better in natural sciences than in human and social sciences. Thus subject composition cannot be disregarded when comparing individual institutions.

EUMIDA provides some useful information in this respect; Data Collection 1 includes at least the indication of which fields are present in each institution, while in Data Collection 2 disaggregation of students and staff data is requested (but not available from all countries). In the best case, separate analysis should be performed for each subject domain; alternative strategies when complete data is not available are restricting the analysis to similar groups of institutions – e.g. only generalist universities or only the technical universities – and excluding those domains for which reliable data is not available (e.g. computing publication to staff data by excluding staff in social sciences and humanities).

Private vs public universities

In the European context, universities are predominantly public institutions. Nevertheless a number of private institutions exist, usually recognized and sometimes partially funded by the government. The classical solution of a dummy variable is appropriate here. It allows estimating the variability in efficiency across the two categories, as well as the variability internal to the categories. Another possibility may be to carry out two separate analyses, one for private and the other for public universities. We notice also that data is not fully comparable between the two categories, owing to different legal status (for example commercial accounting in private universities) and different requirements to collect data.

Age and structure of universities

In other cases, there might be large heterogeneity according to the age of universities, if in the history of higher education a discontinuity has been produced (Bonaccorsi and Daraio 2007b). The age of the university is not what matters per se, but the age is likely to reflect specific characteristics of the institution. Many older universities are likely to be broad research-oriented universities, while the younger ones might be more specialized. This again might affect funding possibilities and research intensity. An interesting case is United Kingdom, where universities include also the so-called "new universities", being the older Polytechnics, which were where transformed into universities in 1992 (with the right of

granting PhD's). A methodological solution for efficiency analysis is to include the age of the university as a descriptor and to consider it in estimations, by using the two age variables included in the database. Another solution, once the qualitative analysis has shown large enough differences, is to introduce a break in the sample and make estimates separately for the age categories.

5.6 Using indicators as tools for societal and scholarly debate

The reader of this Chapter might have been a bit disappointed by discovering the range of comparability problems that affect the EUMIDA dataset. Moreover, it is clear that while some improvements can be made on the data and their collection procedures, a number of comparability problems are so deeply entrenched in national and institutional specificities that any solution can hardly be foreseen. And some of the comparability problems relate simply to the fact that trying to compare a college in Estonia with an ancient university in Portugal might not make much sense, given the diversity of their national context, activities and history.

We don't think this is a right attitude. Very much like other institutional databases, the EUMIDA dataset is a powerful tool to analyze the functioning of European higher education and to provide a view of its characteristics across about 30 national systems and 2,500 different institutions. However, when exploiting it, first, care must be taken to understand the limitations and to check for comparability problems (hence metadata is an essential component of the dataset) and, second, some exploitation strategies and indicators are more robust than others, as shown in the discussion of comparability problems in Section 3 above.

At the end, however, indicators are not objective answers, but constructs based on specific representations and policy choices and these lead exactly to the simplifications which are required to produced usable indicators; for example, one might decide that there is no objective reason why levels of staffing should systematically differ between subject domains and thus disregard subject composition when calculating staff to students ratios; or, alternatively, consider that having 50% more staff per student in natural sciences is the "right" level and thus compensate accordingly for subject differences when comparing institutions.

This means that indicators are not answers to questions, but contributions to the political and social debate; other actors might question them on ground of their technical quality, but especially of their underlying assumptions. When comparing higher education institutions, figures from EUMIDA database should not be used directly to take decisions, but as tools to raise questions and to nurture the debate among policy actors. This implies also that, besides improving the data quality, the creation of suitable arenas for debating these results

– like conferences, forums or expert groups – should deserve high priority in higher education policy.

5.7 Coverage of data

One of the methodological challenges of the EUMIDA feasibility study was to build up a census of institutions of higher education whose perimeter was not distant from the one emerging from official statistics of students. As it is well known, statistical offices follow the UNESCO-OECD-Eurostat Manual (UOE 2006) in collecting data on students at the level of higher education, labelled ISCED 5 and ISCED 6. This data registers students attending courses of a given duration, independently on the nature of the institution delivering these courses. The UOE statistical approach is program-oriented, while the EUMIDA study was aimed at building an institution-oriented approach. In most cases the two perspectives broadly coincide, like for example in universities. But there are cases of institutions whose main missions and activities are outside the educational sector but to a limited extent deliver also tertiary education diplomas, usually of vocational training type. These providers are typically small organizations, affiliated to professional associations or local institutions, and deliver accredited courses without creating self-sustaining educational institutions. Consequently, the students of these courses will be registered as ISCED 5b students, but the providers of the courses do not qualify as "institution" in the sense of EUMIDA. Another difficulty comes from small-scale educational providers, which cannot be considered as stand-alone institutions, but are just departments of other units or parts of the public administration.

It is then important to examine the coverage of EUMIDA data with respect to student statistics. Table 43 summarizes the situation as for the year 2008.

The institutions registered in the EUMIDA census enrol 15,528,151 students, or about 90% of the total number of UOE students. We are therefore 'missing' 1,515,849 students, or less than 10% of the UOE total. This seems to be a reasonable coverage, taking into account that what is left outside the picture is likely to be fragmented into a large number of small providers of professional courses. Furthermore, the difference is largely explained by three national cases, all related to large countries:

- In Germany (difference: 219,793 students), vocational training at tertiary level is not considered part of the educational system and has been left outside the perimeter (a similar situation applies to Switzerland with a difference of 51,905 students and Belgium with as many as 141,300);
- In Spain, the entire private sector has been left outside the perimeter (difference: 609,599 students, most in the professional tertiary education sector);
- In Poland, there is a large but fragmented professional education sector, which has been only partially covered in terms of institutions: while the total number of

institutions in the perimeter is very large, at 457, still data leave a difference of 211,531 students.

Table 44. Coverage of EUMIDA census as a proportion of UOE ISCED 5 and 6 students. Year 2008

Country	Number of institutions in DC 1	Number of students in DC1 (2008)	Number of students ISCED 5-6 UOE (2008)	Coverage (%)	Difference (number of students)
Austria	68	270.131	284.800	94,85	14.669
Belgium	87	260.400	401.700	64,82	141.300
Bulgaria *	59	259.945	264.500	98.27	4.555
Cyprus	37	23.083	25.700	89,82	2.617
Czech Republic	73	379.220	392.500	96,62	13.280
Denmark ***	n.a.	n.a.	230.700	n.a.	n.a.
Estonia	34	68.168	68.200	99,95	32
Finland	49	296.569	309.600	95,79	13.031
France ***	n.a.	n.a.	2.164.500	n.a.	n.a.
Germany	410	2.025.307	2.245.100	90,21	219.793
Greece **	60	584.274	602.900	96,91	18.626
Hungary	72	381.033	413.700	92,10	32.667
Italy	243	2.005.240	2.013.900	99,57	8.660
Ireland	21	155.036	178.500	86,85	23.464
Latvia *	61	112.567	127.800	88,08	15.233
Lithuania	46	202.509	204.800	98,88	2.291
Luxembourg	1	3.526	3.000	117,53	-526
Malta	4	10.515	9.500	110,68	-1.015
Netherlands	59	629.551	602.300	104,52	-27.251
Norway	68	215.851	212.700	101,48	-3.151
Poland	457	1.954.469	2.166.000	90,23	211.531
Portugal	138	373.002	376.900	98,97	3.898
Romania	82	1.034.214	1.056.600	97.88	22.386
Slovakia	33	232.936	229.500	101.49	-3.436
Slovenia	13	96.323	115.400	83,47	19.077
Spain **	47	1.167.901	1.777.500	65,70	609.599
Sweden	49	365.162	406.900	89,74	41.738
Switzerland	36	172.595	224.500	76,88	51.905
United Kingdom	150	2.248.624	2.329.500	96.52	80.876
Total ***		15.528.151	17.274.700	89.88	1.515.849

* EUMIDA data refer to 2009

** EUMIDA and EUROSTAT data refer to 2007

*** France and Denmark not included (based on the information from experts in the country report, the coverage for France would be 76.22% and 87.79 % for Denmark)

These national cases (Germany, Belgium, Poland, Spain and Switzerland), well explained by institutional discrepancies in the coverage of providers of ISCED 5b courses, cover 1.234.177 students, or 79,4% of the total difference.

This means that the differences with respect to EUROSTAT data do not depend on mistakes or discrepancies between aggregate data and microdata. Rather, they depend mainly from a deliberate choice of Statistical Authorities to exclude very small schools, delivering ISCED 5b courses, from the perimeter of higher education institutions as defined in the EUMIDA project. This should not be a concern, since the focus of EUMIDA is on institutions, not courses. The perimeter of institutions will inevitably cover a subset of the aggregate number of students, since some of them are enrolled in courses that are not delivered by stable institutions. Rather, it is remarkable that only 9% of aggregate students are left out of the perimeter. It means that what is left out is perhaps a large number of course-delivering units, none of which has the structural characteristics to be considered an institution.

At the same time, for a few countries the number of students registered in EUMIDA exceeds the Eurostat data for the same year, with no clear explanation.

Here the coverage exceeds 100%, i.e. Eurostat aggregate data record less students than the sum of EUMIDA institutions. This seems to violate our own methodological definitions, according to which 'The perimeter for EUMIDA data collection is by definition smaller than the perimeter of UOE data collection' (see above 2.3.1). Looking at the data, these problems come from two very small countries with just one institution (Luxembourg and Malta) and two other countries (Netherlands and Norway). Noting that Eurostat data is rounded to the hundred units, we interpret this discrepancy as a minor measurement error, as if for the institutions in the EUMIDA dataset the number of students might have included a few units in addition to, (or perhaps a few days after?) the aggregate number recorded in the form sent to Eurostat.

Taking into account these elements, it can be said that the data covers the overwhelming majority of higher education in Europe.

6. Characterization of the higher education landscape

The European higher education landscape is characterized by a high degree of diversity with respect to period of foundation, institutional type, size, mission and profile. This Chapter is aimed at describing that diversity according to different dimensions.

The EUMIDA census covers 2,457 institutions in 27 European countries (all EU minus France and Denmark due to data availability problems, plus Norway and Switzerland). The estimated number of institutions including Denmark and France is approximately 2,900.

The European landscape of HEIs can broadly be distinguished in two groups: research active and non-research active HEIs. The data collection process in EUMIDA followed this broad distinction and includes 1,364 research-active HEIs and 1,093 non research-active HEIs. For all HEIs a core set of data, allowing a broader characterization of higher education institutions across Europe, were collected. The following analysis is based on this dataset. Hence it includes data on slightly less than 2,500 HEIs (as long as the information is available) but can rely on a limited set of indicators.

Data that is "confidential" is not taken into consideration for the analysis. Due to comparability requirements, available data is not included in the analysis when it was flagged "deviates from definition". However, data that was flagged "not applicable", "provisional value" or "unreliable or uncertain data" is considered, in order to provide a characterization of the European higher education landscape as broad as possible.

6.1 Historic development of HEIs in Europe

The European higher education landscape is characterized by a long tradition. The most ancient university in Europe, hence in the world, i.e. the Università degli Studi di Bologna, was established in 1088. Of the ten most ancient universities in Europe (that are still existing), seven are located in Italy (Università degli Studi "Federico II", Università degli Studi "La Sapienza", Università degli Studi di Bologna, Università degli Studi di Genova, Università degli Studi di Padova, Università degli Studi di Perugia, Università degli Studi di Pisa), one in the Czech Republic (Univerzita Karlova v Praze), one in Austria (Universität Wien) and one in Poland (Uniwersytet Jagiellonski w Krakowie). All of these are public, research active and have the right to award doctorates.

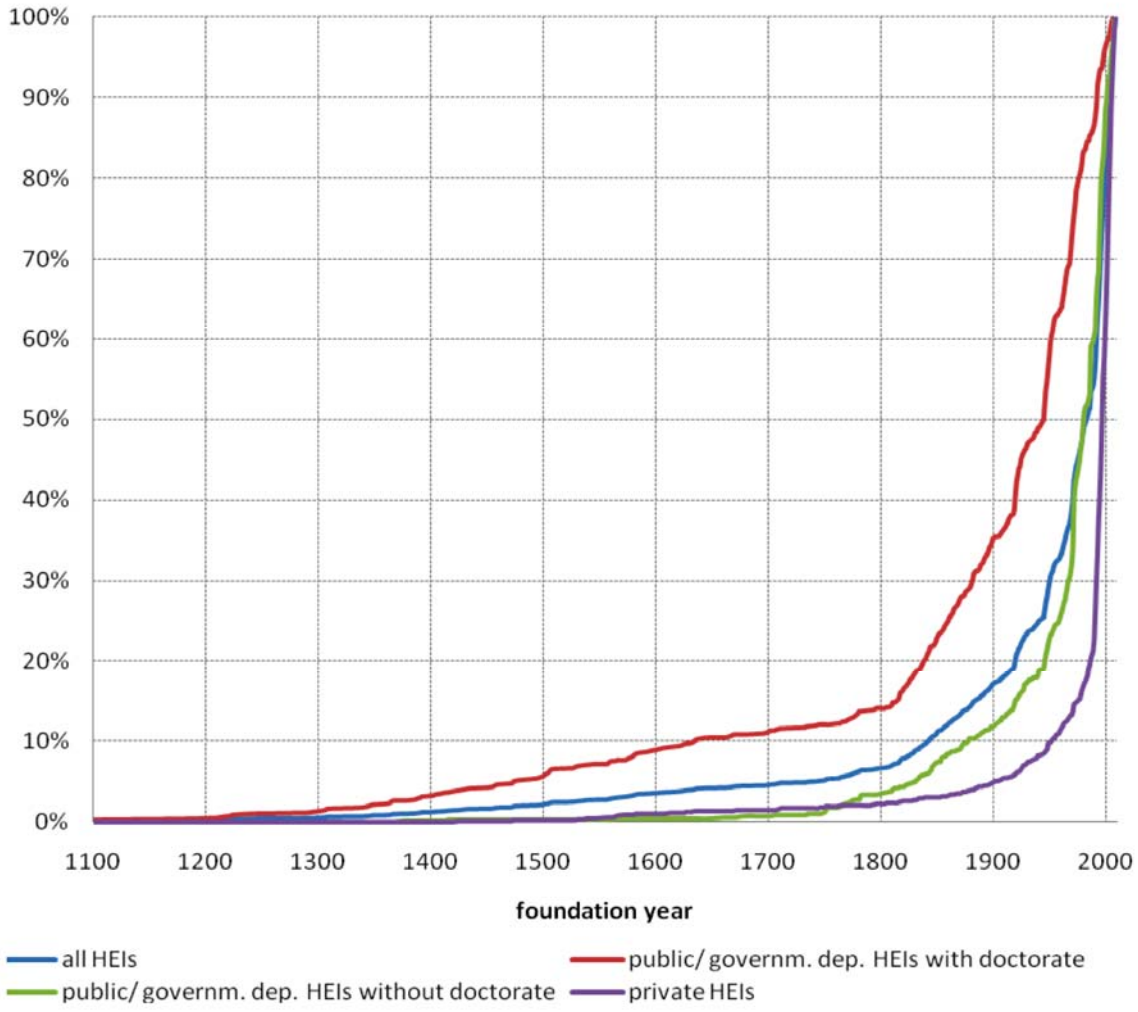


Figure 8. Historic development of the European higher education landscape by category
Source: EUMIDA. Dataset does not include Denmark and France (no data available).

It is useful to examine the process of creation of HEIs by sub-groups (see Figure 8). Public (respectively, government dependent) HEIs with the right to award doctorates are on average older than other types of HEIs. At the beginning of the 20th century, more than one third of these HEIs were already in place. However, half of them have been established after World War II. In comparison, private HEIs were almost non-existent before the 20th century, since 95% of them were established after 1900. The take off of public institutions without doctorate started in early 19th century.

Although HEIs have a very long tradition in Europe, the increasing need for higher education lead – and still leads - to the establishment of new HEIs (see Figure 9). Roughly 45% of the European HEIs have been established since 1990. This ongoing growth is primarily based on private HEIs and to a lesser extent on public (respectively government dependent) HEIs without the right to award doctorates. In terms of the number of institutions, private HEIs became a major part of the European higher education landscape in the past decades only.

Almost 80% of private HEIs in Europe have been founded in the past two decades (see Figure 9).

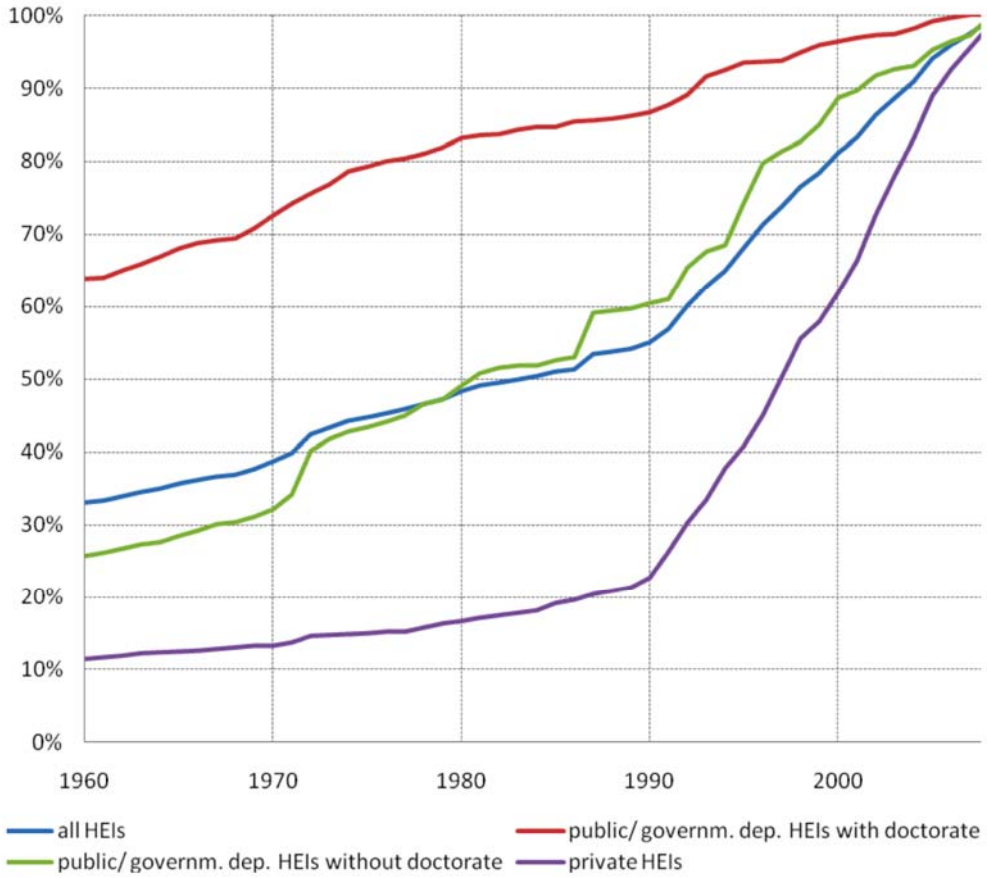


Figure 9. Development of the European higher education landscape in the past 50 years by category

Source: EUMIDA. Dataset does not include Denmark and France (no data available).

The analysis of the creation of universities by cohort offers several illuminating insights.

Overall, 6.6% of HEIs have been founded in 1800 or before, 10.5% from 1801 to 1900, 9.3% from 1901 to 1945, 13.5% from 1946 to 1970, 17.0% from 1971 to 1990, 26.3% from 1991 to 2000 and 16.8% from 2001 to 2009.

However, there is significant variability across different types of HEIs (see Figure 10, top panel): public (respectively government dependent) HEIs with the right to award doctorates show a much larger share of old HEIs than other types of HEIs. A similar pattern is visible for research-active HEIs and for large (15,000 to 30,000 students enrolled) and very large HEIs (more than 30,000 students enrolled). National systems characterized by a larger share of old institutions do not exhibit a clear pattern of regional (or supra-national) clustering. Countries that are characterised by a high share of old HEIs include Austria, Germany, Hungary, Ireland, Italy, Romania, Sweden, and the UK - and hence Anglosaxon countries, countries with a Humboldtian tradition, Eastern European countries and a Mediterranean

country. In contrast, the share of private and small HEIs that are young (founded in the past two decades) is much larger than average in some countries. These include Austria, Belgium, the Czech Republic, Latvia, Poland and Slovenia. This highlights the changes in higher education systems that have occurred in many former socialist countries, leading to the establishment of new HEIs, many of which private. However, this also shows that the higher education landscape of other European countries changed considerably in the past two decades. Interestingly, the Austrian higher education system is characterised by an above-average share of both old and young HEIs, highlighting the rapid change of a very old and traditional higher education system in the last two decades.

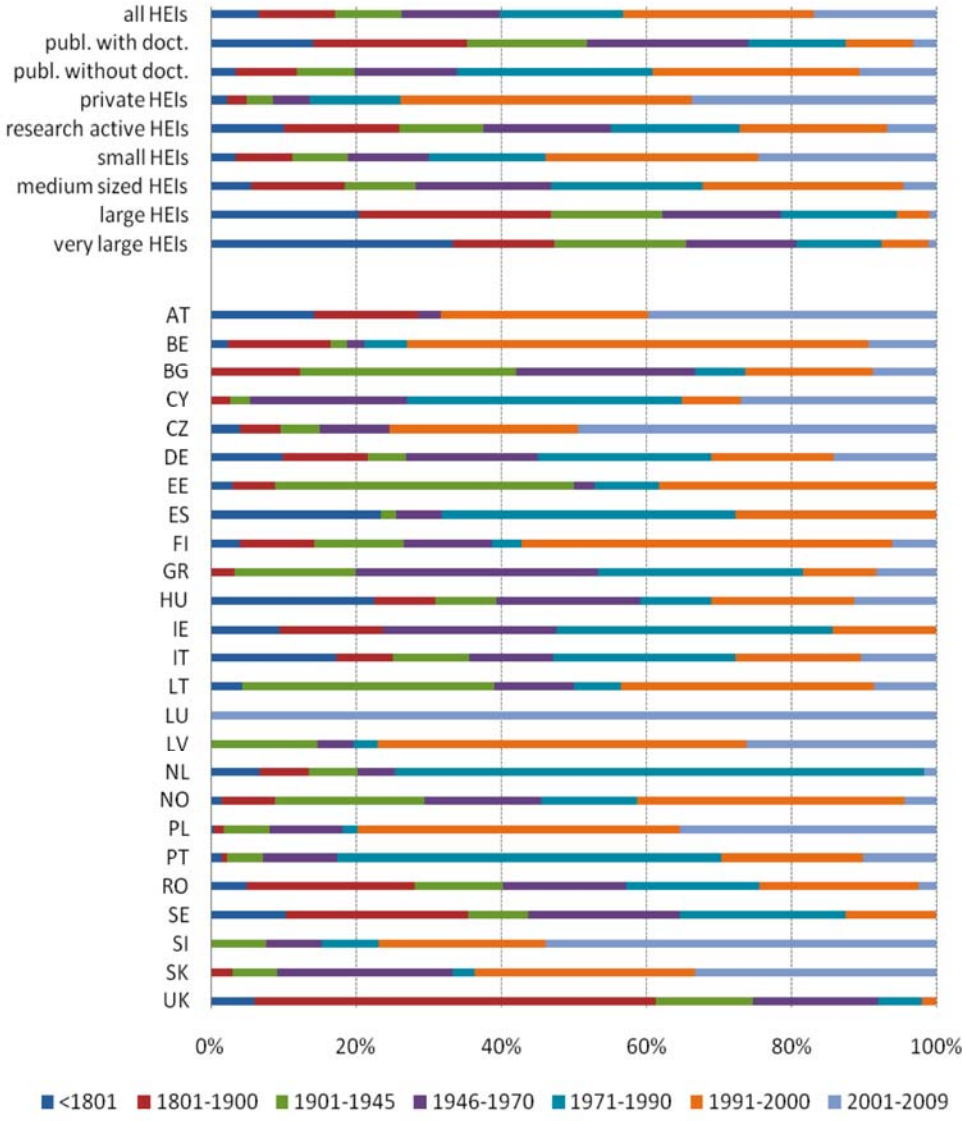


Figure 10. Age distribution of the European higher education institutions by category and country

Source: EUMIDA. Dataset does not include Denmark and France (no data available) and Malta and Switzerland (limited data availability)

Data shown in Figure 10 suggests that the European higher education landscape has not yet stopped growing in terms of absolute number of institutions created. These are mostly concentrated in the segments of non-doctorate HEIs and of private institutions. Also, slightly less than 30% of research-active institutions have been created after 1990.
However, since younger HEIs are on average much smaller than older HEIs when the HEIs are weighted by the total number of students, the importance of older and more traditional HEIs becomes apparent. Although almost 45% of European HEIs have been founded after 1990, these account for 20% of students only. This is the same share as very old HEIs (founded before the 19th century). One third of students is enrolled in HEIs that have been founded before 1900 and two thirds in HEIs that have been founded before 1971. Hence, although the massification of higher education has led – and still leads – to the establishment of new HEIs, traditional universities absorb a relatively large share of additional students.

6.2 Size of student body

The size distribution of European HEIs is very uneven. While there is large number of very small HEIs, there are also some HEIs with more than 100.000 students enrolled (aggregated number of students ISCED 5 and ISCED 6): These are the Universitatea "Spiru Haret" Bucuresti, located in Romania, The Open University, located in the UK, both distance education institutions, and Università degli Studi "La Sapienza", located in Italy. On the other hand, there are universities, active in teaching, with less than 10 students. These include Seminarium Kościoła Starokatolickiego "Mariawici" w Płocku, Wyższe Seminarium Duchowne Kościoła Polsko-Katolickiego w Warszawie, Wyższe Seminarium im. Jana Łaskiego (Metodyści) w Warszawie located in Poland, Latvijas Nacionālā Aizsardzības akadēmija located in Latvia, Vysoká škola cestovního ruchu a teritoriálních studií v Praze, spol. s r.o. located in the Czech Republic, ISS International Business School of Service Management Hamburg located in Germany, Instituto Superior de Psicologia Aplicada de Beja located in Portugal and Istituto musicale Merulo located in Italy. Considering the entire European higher education landscape, the median higher education institution has slightly more than 6,500 enrolled students (aggregated number of students ISCED 5 and ISCED 6), while the average value is less than 2,000. Considering the entire distribution (see Figure 11) it appears that 20% of students are enrolled in less than 2.5% of institutions and 50% of students in 10% of the largest HEIs. At the other extreme of the distribution, 20% of smallest HEIs account for 0.5% of the total number of students only and half of the HEIs account for slightly more than 4.5% of students enrolled. The 100 largest HEIs in Europe account for 30% of the students and the largest 200 HEIs for almost 50% of students.

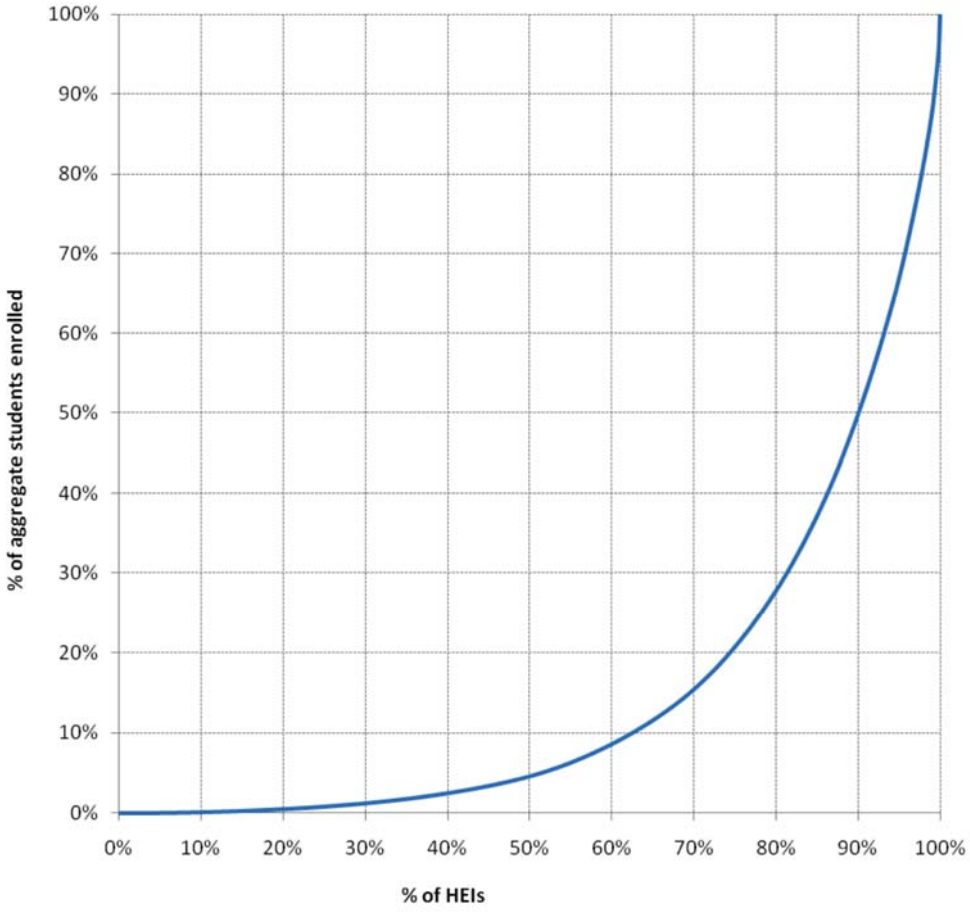


Figure 11. Size distribution of European higher education landscape

Source: EUMIDA. Dataset does not include Denmark and France (no data available).

It is useful to carry out the analysis of size distribution by introducing size categories. There are several possible size categories (see Daraio et al., 2010). In the EUMIDA project the size categories suggested by the U-Map project (see van Vught, F.A. et al., 2010) have been adopted, as follows:

- Very large: more than 30,000 students enrolled
- Large: between 15,000 and 30,000 students enrolled
- Medium sized: between 5,000 and 15,000 students enrolled
- Small: less than 5,000 students enrolled.

Following this classification, two thirds of European HEIs are small, 18.5% medium sized, 10% large and 4% very large. Considering different categories, the size of the university student body is a function of age (young universities are all very small or small), highest degree delivered (doctorate awarding HEIs are larger) and research activity (research active HEIs are larger). With respect to the year of foundation, the share of small institutions decreases monotonically with age.

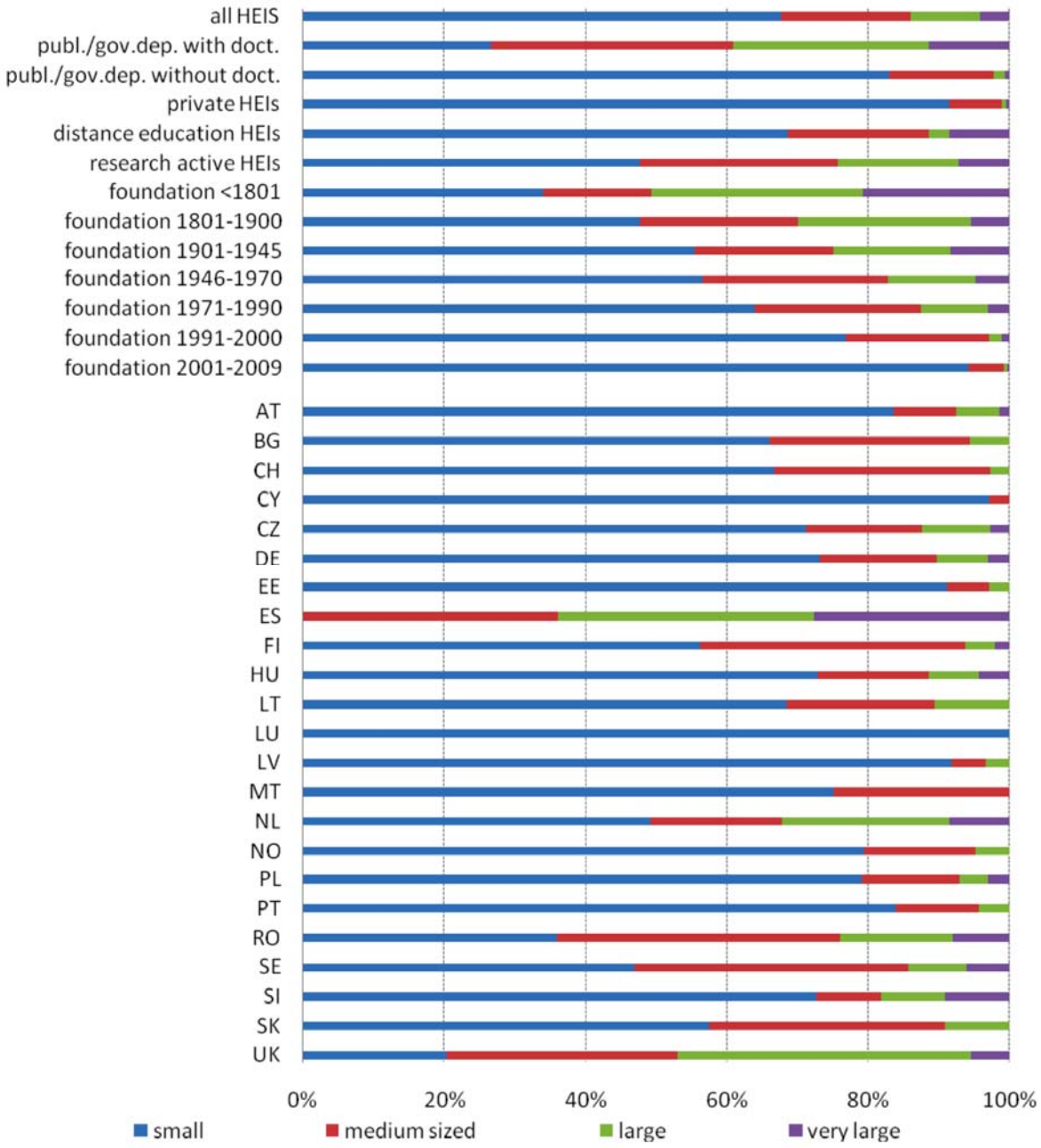


Figure 12. Size distribution of the European higher education institutions by category and country

Source: EUMIDA. Dataset does not include Denmark and France (no data available) and Italy (data on younger HEIs not available).

Interestingly, there are national differences in the size distribution of the student body of HEIs, but only to a limited extent, as the median institution is quite similar across many countries. Two important exceptions are UK and Spain[22] (see Figure 12). Italy is another exception, but is not reported here due to lack of data on younger institutions.

[22] For Spain the limited perimeter might explain the situation (with no small HEI in the dataset).

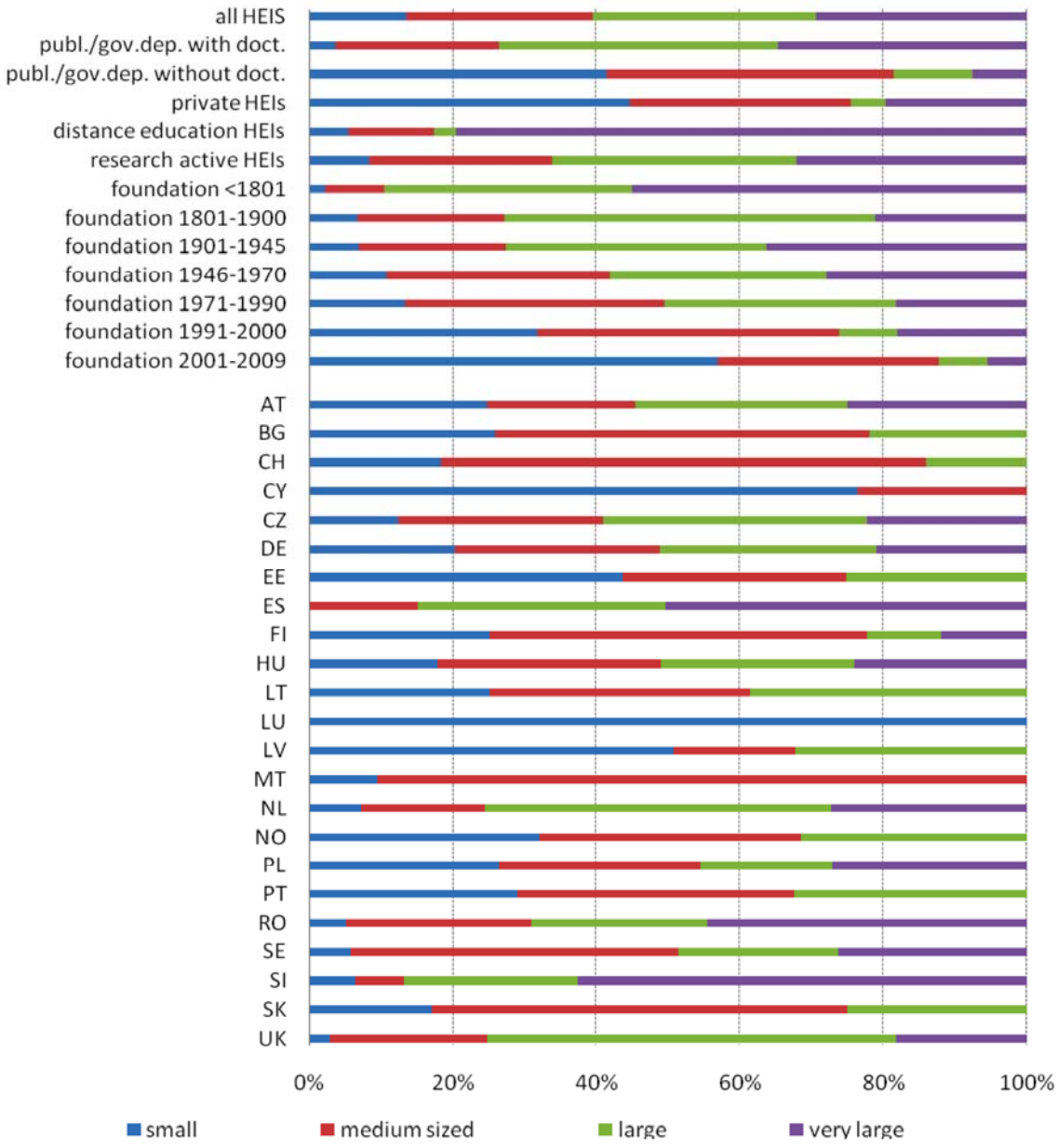


Figure 13. Size distribution of the European higher education institutions (weighted by # of students) by category and country

Source: EUMIDA. Dataset does not include Denmark, France (no data available), Belgium, Greece, Ireland, Italy (limited data availability).

A very different picture emerges when HEIs are weighted by the size of the student body (see Figure 13). Although two thirds of the European HEIs are small, only 13,5% of students are studying at small HEIs. Only for very young HEIs (founded past 2000), small HEIs account for the majority of students. To the contrary, 60% of students study at large and

very large HEIs. Countries such as Austria, the Czech Republic, Germany, Estonia, Hungary, the Netherlands, Romania, Slovenia and the UK are dominated by large and very large HEIs. For distance education HEIs, 80% of students are enrolled in very large institutions, highlighting the role of a small number of very large institutions (Universitatea "Spiru Haret" Bucuresti, located in Romania, The Open University, located in the UK, and Fernuniversität Hagen, located in Germany).

In addition to the overall size of the student body, it is also interesting to consider the differences in focus on ISCED 5 and ISCED 6 students (see Figure 14). If the number of students at level ISCED 6 is considered, then some very large universities exist, with more than 5,000 students. These are Universidad Complutense de Madrid, located in Spain, Universität Wien, located in Austria, Univerzita Karlova v Praze, located in the Czech Republic, Universität zu Köln, located in Germany and Helsingin yliopisto, located in Finland. Hence universities with a very large number of students at level ISCED 6 can be found in many different countries.

Although these HEIs are characterised by a very large number of students at level ISCED 6, they are not particularly focusing on PhD education. When considering the focus on students level ISCED 6 it is possible to differentiate HEIs only offering programmes at level ISCED 6 (e.g. Българска академия на науките and Селскостопанска академия, located in Bulgaria, Istituto Italiano di Scienze Umane, Scuola IMT Alti Studi and Scuola Normale Superiore, located in Italy and Academie Des Beaux-Arts Et Des Arts Decoratifs De Tournai located in Belgium). On the other hand, there are HEIs with students at level ISCED 5 and ISCED 6 that are characterised by the high share of ISCED 6 students. Among the largest universities with a share of ISCED 5 students over ISCED 6 students of less than 5 are the University of Oxford, the University of Cambridge, University College London, Imperial College of Science, Technology and Medicine, located in the UK, Teknillinen korkeakoulu, located in Finland, Eidgenössische Technische Hoschule Zürich and Universität Basel, located in Switzerland. In addition there are some smaller HEIs with a strong focus on education at doctoral level, very often with a specific mission (e.g. Theological HEIs in Germany: Philosophisch-Theologische Hochschule Frankfurt a.M., Philosophisch-Theologische Hochschule St. Augustin, Theologische Fakultät Paderborn, Theologische Hochschule Vallendar, Theologische Fakultät Fulda, or medical resp. veterinary institutions like Private Universität für Gesundheitswissenschaften, Medizinische Information und Technik (UMIT), located in Austria, Tierärztliche Hochschule Hannover, located in Germany, Norges veterinærhøgskole, located in Norway, or the Institute of Cancer Research and the London School of Hygiene and Tropical Medicine, located in the UK).

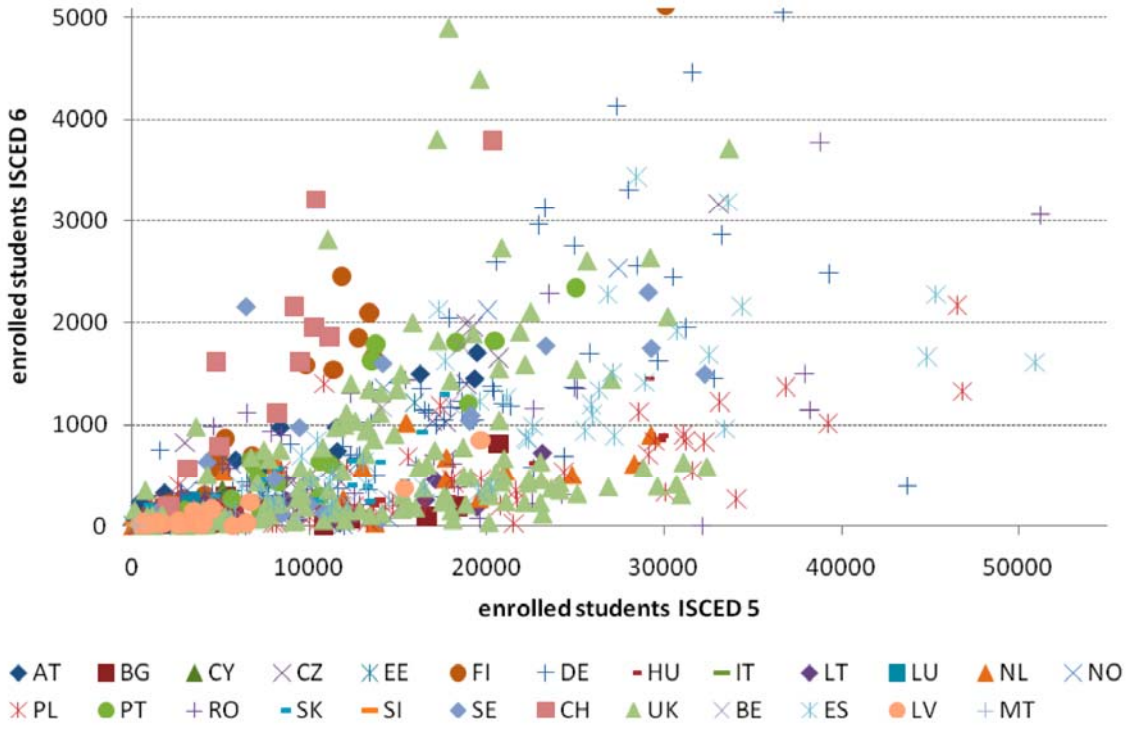


Figure 14. Enrolled students ISCED 5 and ISCED 6 by country

Source: EUMIDA. Dataset does not include Denmark and France (no data available), Belgium, Greece, and Ireland (limited data availability)

6.3 Legal status

The European higher education landscape is dominated by public institutions: More than half the European HEIs are public, another 5% are government dependent private, i.e. have a private governance but the public sector dominates in the contribution to the budget. However, in the past two decades private HEIs gained importance. Three quarters of the existing HEIs founded after 2000 and more than half founded between 1991 and 2000 are indeed private – and the majority of these are small (see Figure 15).

The size of HEIs differs considerably by legal status: On average the student body of public HEIs is twice as large as that of government dependent private HEIs and almost five times as large as that of private HEIs.

National higher education systems show large differences in the relative importance of HEIs by legal status. In Ireland and Luxembourg only public HEIs exist[23]. In addition, in Switzerland, Greece and the UK more than 80% of the HEIs are public. Government dependent private HEIs are of considerable importance only in a small number of countries: Austria, Belgium, Finland, Netherlands, Norway and Slovenia. On the other hand, the higher education systems of Cyprus, Czech Republic, Hungary, Poland and Portugal are dominated

[23] The Spanish perimeter includes public HEIs only, although a considerable number of private HEIs do exist.

in number by private HEIs (see Figure 15). However, if the HEIs are weighted by the student body, then this does no more hold true, apart from Cyprus (Figure 16). The Netherlands is the only other country that is not dominated by public HEIs, weighted by the student body, as Hogescholen are mainly government dependent private.

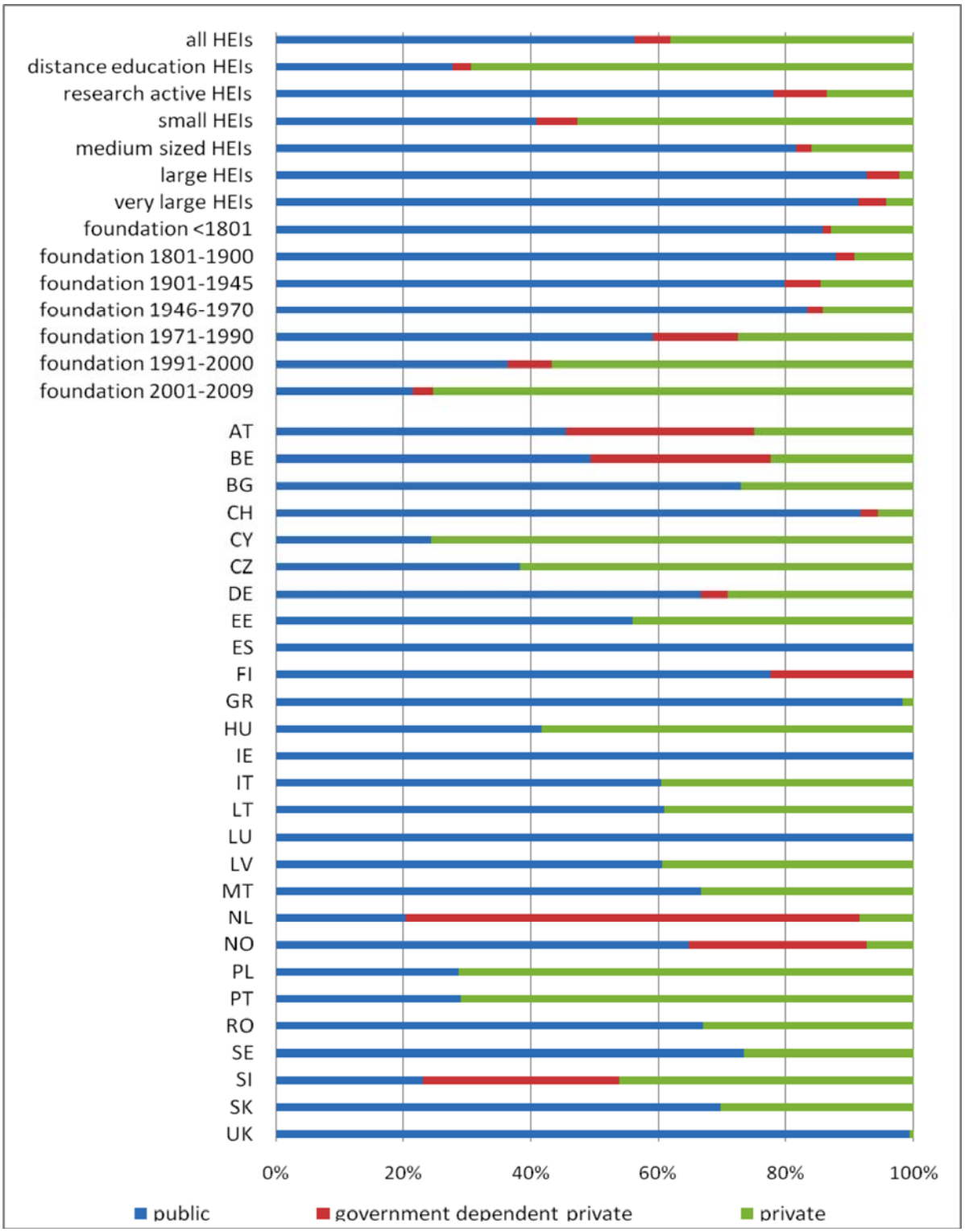


Figure 15. Legal status of the European higher education institutions by category (number of institutions) and country

Source: EUMIDA. Dataset does not include Denmark and France (no data available).

Overall the share of students in public HEIs is more than 80%. This is also the case for the majority of national higher education systems. Countries with more than 20% of students in private HEIs are Bulgaria, Latvia, Malta, Poland, Portugal and Romania. Hence private HEIs play an important role primarily in Eastern and Southern European countries, where a massive expansion of the higher education took place in the past two decades.

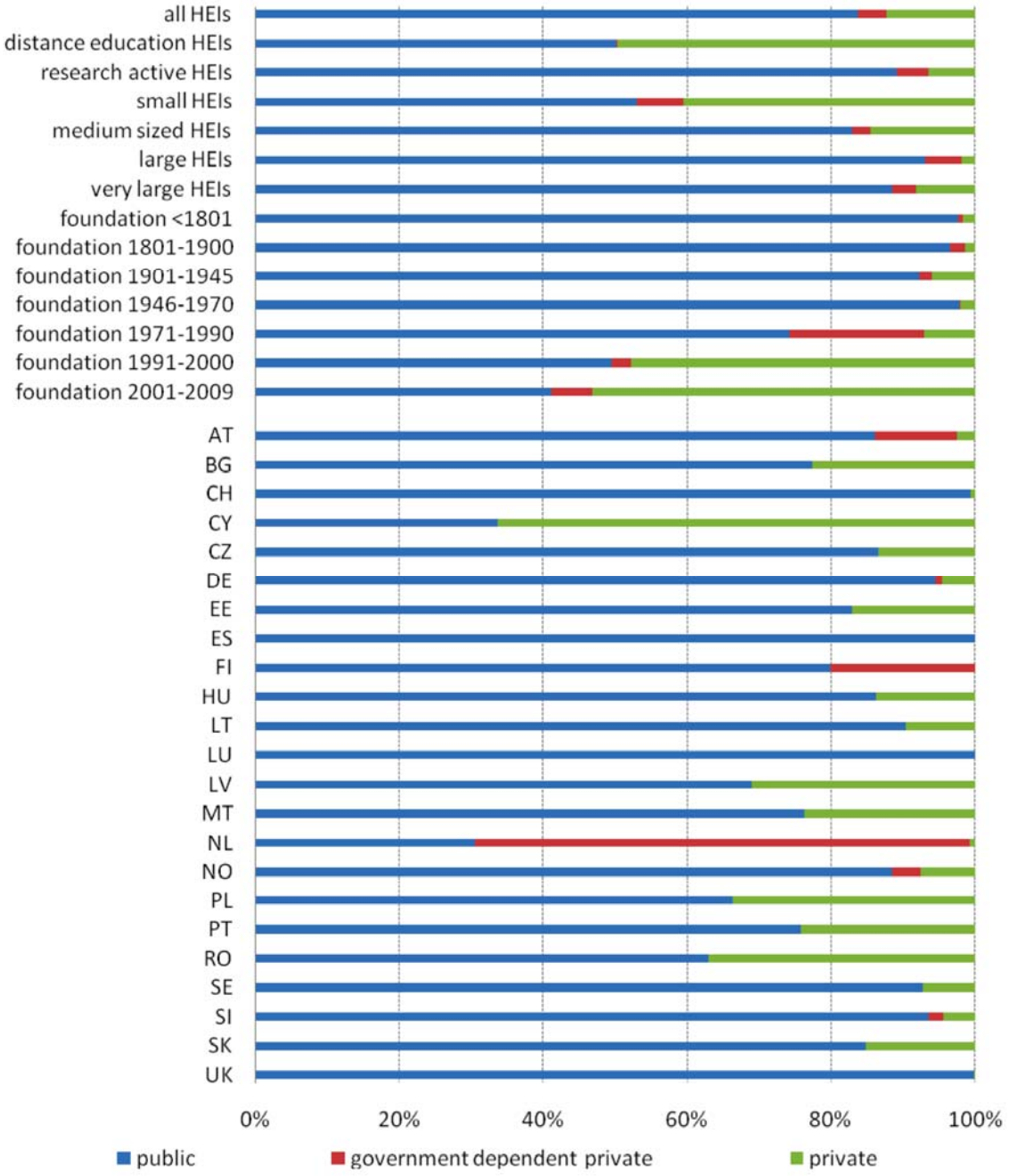


Figure 16. Legal status of the European higher education institutions (weighted by student body) by category and country

Source: EUMIDA. Dataset does not include Denmark and France (no data available); Belgium, Greece, Ireland, Italy (limited data availability).

6.4 Highest degree delivered

In a few cases, there are missing data. Of the institutions for which information is available (n=2,410), 846 deliver only Bachelor degrees, 672 deliver degrees up to Master or equivalent degrees pre-Bologna process, while 892 deliver degrees up to Doctorate.

The EUMIDA methodology developed a set of criteria for classifying institutions in terms of their actual research activity[24].

First, the higher education landscape is composed by three groups that are almost equivalent in number, delivering degrees at bachelor, master or PhD level. The system is not apparently organized as a pyramid, with a large base of institutions covering lower level curricula, but as a clepsydra.

Second, institutions of higher education specialised in bachelor degrees (n= 846) include two groups. The largest group includes generalist institutions, delivering bachelor degrees across many fields. The most important example is the model of Fachhoschule, which in some countries have been labelled University of Applied Sciences. Other examples include University colleges, or Other higher education institutions. The label "college" is used only by 129 institutions. The remaining group includes institutions delivering degrees in highly specialised fields, such as Art and drama, Languages, Theology, or Public Administration.

Third, while the number of institutions is similar in the three groups, the overwhelming majority of students are enrolled in institutions delivering up to the doctoral degree, or traditional university model institutions. Almost 80% of students are enrolled in institutions with the right to award the doctorate, while 8.8% are enrolled in the 846 institutions that deliver only bachelor degrees and 12.6% are enrolled in institutions that deliver up to master degrees. Although the distribution of institutions in the three groups is balanced, almost 80% of European students go to a university-like institution, even if they want to attend courses that will not lead to the highest degree that the university can confer.

[24] The following definitions were offered in the Handbook:
"Among the whole population of higher education institutions, we distinguish the research-active ones, i.e. those having an institutionalised research activity. This distinction is relevant because of the specific functions and organizations of these institutions.
The definition of research active does not imply a specific level of *research intensity* and care should be taken in distinguishing between research-active and research-intensive institutions (exceeding some threshold, like the one used in the Carnegie classification). However, it implies that research is considered as constitutive part of institutional activities and is organised institutionally and with a durable perspective. Criteria for inclusion are then the following:

- Existence of an official research mandate.
- Existence of research units institutionally recognised (for example on the institutional website).
- Inclusion in the R&D statistics (availability of R&D expenditure data), as sign of institutionalised research activity.
- Awarding doctorates or other ISCED 6 degrees.
- Consideration of research in institutions strategic objectives and plans.
- Regular funding for research projects either from public agencies or from private companies.

Institutions fulfilling at least three of these criteria should be included.
On the contrary, diffused research activities undertaken by teachers on their own interest are not a sufficient criterion to consider an institution as research-active".

Institutions dedicated to bachelor or master degrees, not having the legitimisation of delivering doctorate degrees, attract slightly more than 20% of students.

Fourth, the private educational sector covers 12% of the total number of institutions, of which roughly one quarter in institutions delivering only bachelor degrees, half master, and another quarter doctorate. However, given the concentration of students in the latter category, the private sector covers more than 30% of the students enrolled in institutions offering only bachelor, 47% of those whose institution offers up to the master degree, but only slightly more than 3% of those that attend in institutions offering up to the doctorate. It seems clear that the private sector addressed those educational needs left unanswered by the public sector, particularly in specialised areas and in those countries in which economic growth is more recent.

Finally, there is no perfect overlapping between the notion of research active institution and the model of institution delivering the doctorate degree. On one hand, there is a handful of institutions that, while legitimated to deliver doctoral degrees, yet are non-research active (n= 40). More importantly, among research active institutions there are institutions that deliver only the bachelor (n= 226) or up to the master degree (n= 317). This group includes several Fachhoschule and Universities of Applied Sciences, as well as many specialised institutions. It accounts for 39% of research active institutions.

The right to award a doctorate is a function of size, age and legal status of HEIs: Whereas the share of small HEIs with the right to award a doctorate is only 17%, this share is more than 90% for large and very large HEIs. And for the youngest group of HEIs (founded between 2001 and 2009) this share is 10% whereas for HEIs founded before 1945 it is above 65%. Only 10% of private HEIs have the right to deliver doctorate degrees, compared to more than half of public (respectively government dependent private) HEIs. However, more than half of the HEIs (40% when weighted by the student body) without the right to deliver a doctorate degree are private. Hence the additional educational demand that was increasingly covered by newly founded private HEIs in the past two decades, seems to be related to undergraduate education.

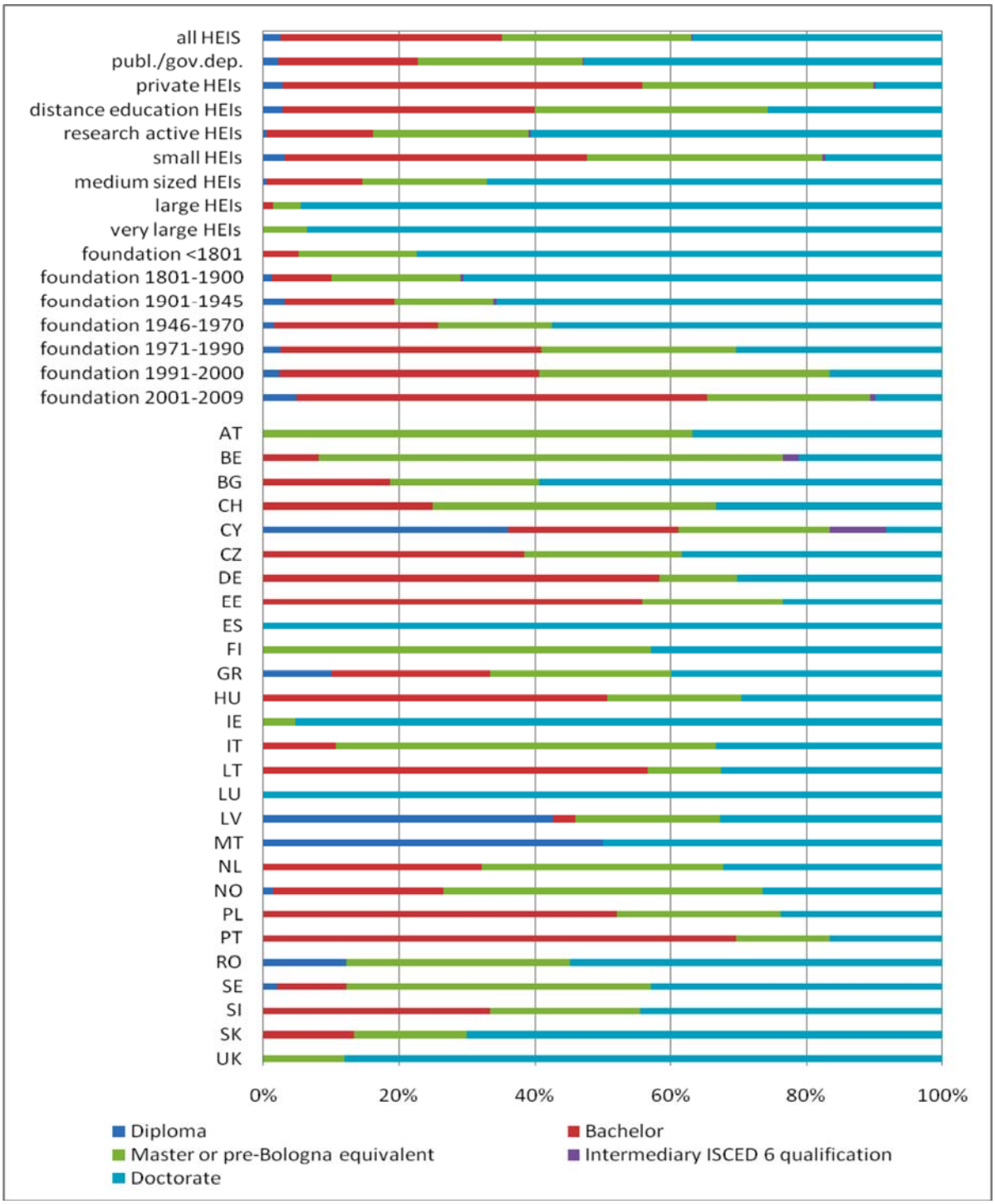


Figure 17. Highest degree delivered by European higher education institutions by category and country

Source: EUMIDA. Dataset does not include Denmark and France (no data available).

It is useful to disentangle the overall picture by examining national patterns. Inspection of data suggests the following:

- Countries in which the share of institutions delivering diploma or bachelor as the highest degree is significant (i.e. above the overall average) include Germany, some of the Eastern European countries (Czech Republic, Estonia, Hungary, Lithuania, Latvia and Poland), and Southern European countries (Cyprus, Malta and Portugal);
- Other countries do not have institutions delivering diploma or bachelor as the highest degree, or exclude them from the statistical definition of higher education: they are Austria, Finland, Ireland, Luxembourg, Spain, and the United Kingdom;
- In other countries the non-doctorate higher education sector is mainly represented by institutions delivering master or pre-Bologna equivalent degrees: they include Austria, Belgium, Finland, Italy, Norway, Sweden and Switzerland; in the case of Italy, the sector is mainly represented by very small specialized institutions in Art (Accademia delle Belle Arti) and Music (Conservatorio), recently attracted in the higher education field;
- Consequently, there are countries in which the proportion of institutions delivering up to the doctorate is overwhelming: they are Spain (100%), although data does not include private universities as already noted previously, Luxembourg with one university, the Anglo-Saxon countries (Ireland, UK), and the remaining group of Eastern European countries (Bulgaria, Romania and Slovakia).

Summing up, we have quantitative and detailed confirmation of national differences in the overall architecture of higher education:

- German speaking countries (Austria, Germany and Switzerland), Belgium, Portugal, Scandinavian countries (Norway, Finland and Sweden) and part of Eastern European countries (Czech Republic, Estonia, Hungary, Lithuania, Latvia, and Poland) have a robust sector of institutions delivering only bachelor and master degrees;
- Latin countries such as Italy have only specialized non-doctorate institutions, or have none, such as Spain; part of Eastern European countries (Bulgaria, Romania and Slovakia) follow this model;
- Anglo-Saxon countries have a small non-doctorate sector, having absorbed the Polytechnic model into the higher education sector (but then they treat them differently de facto through merit-based research funding which goes preferentially to old universities).

Considering not the number of HEIs but the student body, then only the higher education systems of Cyprus and Netherlands are not dominated by HEIs with the right to award doctorates (see Figure 17). Cyprus, Estonia, Germany, Hungary, Lithuania and Portugal have a relatively large sector of institutions delivering only the bachelor, with a share of students exceeding 25%, but apart from Cyprus they have a very small share in the master degree sector.

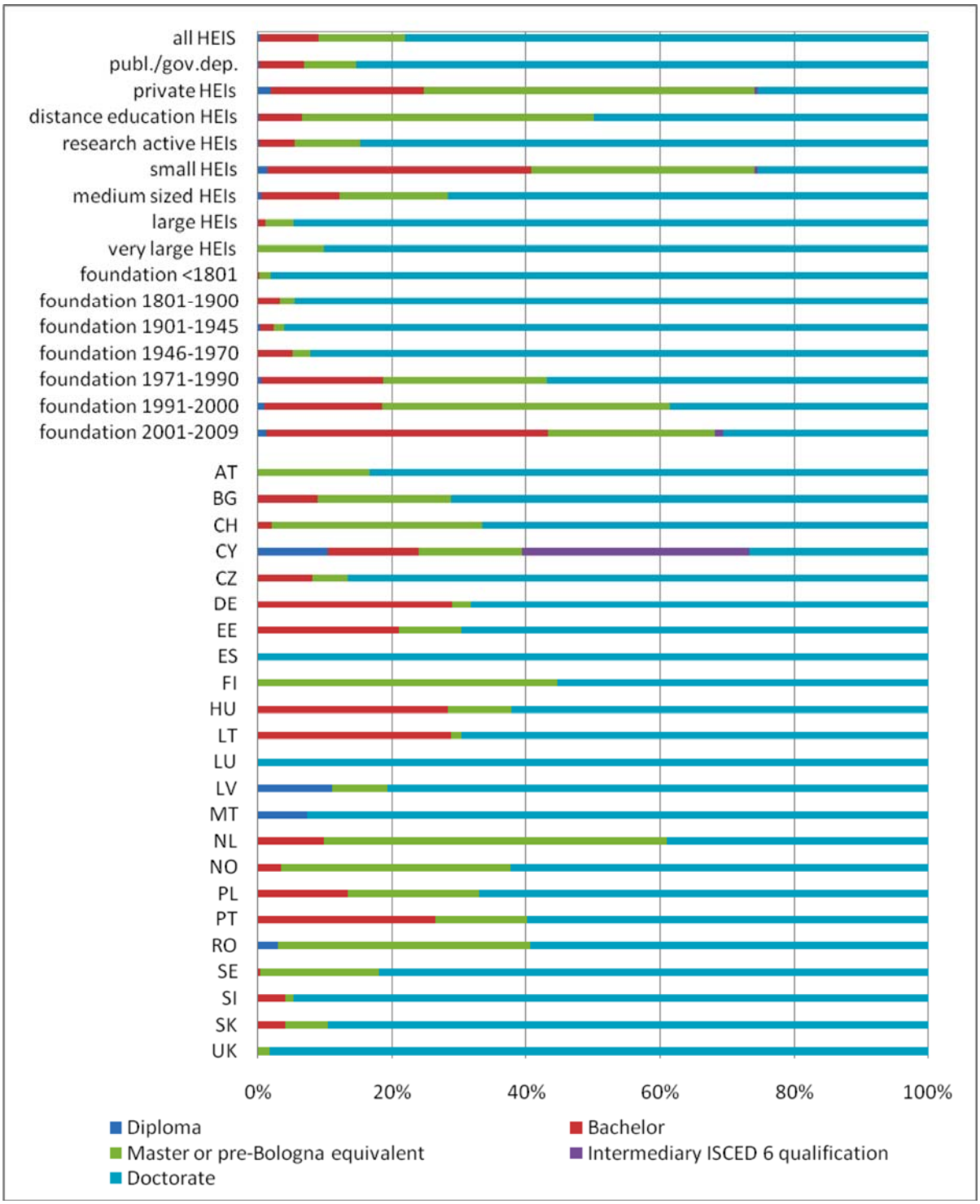


Figure 18. Highest degree delivered by European higher education institutions (weighted by student body) by category and country

Source: EUMIDA. Dataset does not include Denmark and France (no data available); Belgium, Greece, Ireland, Italy (limited data availability).

6.5 Subject mix

HEIs are multi-divisional organizations that offer educational services in one or more domain fields. Depending on their history, the institutional context, and the demand from students, they may have a narrow scope, offering few subject matters, or a broad one.

For purposes of classification, we adopted the U-Map definitions (van Vught et al. 2010), as follows:

- Specialised: 1 to 3 fields of education
- Broad: from 4 to 6 fields of education
- Comprehensive: from 7 to 9 fields of education.

Fields of education that accounted only for a very small share of students (status "present" in the labelling of data) have not been taken into account for the analysis, because they are less representative for the specific profile of a HEI.

Overall almost two thirds of HEIs are specialised, awarding qualifications in less than four fields of education (see Figure 19). The range of subjects differs considerably by type of HEI, however: three quarters of existing HEIs founded in the past two decades and more than 80% of the youngest cohort of HEIs (founded after 2000) as well as 90% of private HEIs are specialised. Again, the range of subject is a function of age, size of the student body and legal status.

Distance education institutions are also an interesting case. More than 80% are specialised. Their focus lies on the field of education "Social sciences, Business and Law", although qualifications in all fields of education are offered by distance education institutions.

As expected, the picture changes greatly when the distribution of institutions is weighted by the number of students. More than 60% of very large HEIs award degrees in a comprehensive range of subjects.

Differences in the range of subjects are also very helpful to illustrate differences in national higher education systems. Countries that are characterised by a high share of comprehensive HEIs (share above 25%) include Germany, Ireland, Malta, Spain, Sweden and the UK.

In contrast, in many Eastern European countries (Bulgaria, Czech Republic, Estonia, Hungary, Poland and Romania) but also in Austria, Cyprus, Italy and Luxembourg, more than 75% of HEIs are specialised. The main reasons for this pattern seem to be the openness of the higher education system to establish private HEIs (that are primarily specialised) or the inclusion of dedicated training institutions (e.g. teacher training) or art schools in the perimeter of higher education.

A closer look at the different fields of education supports this: 45% of HEIs offering degrees in field of education "Humanities and Arts" are specialised (of these more than half offers education only in this field). A similar observation point to Law and Business schools: 45% of HEIs offering degrees in the field of "Social sciences, Business and Law" are specialized.

In contrast, only about 15% of HEIs awarding degrees in the field of "Agriculture" are specialized, while 55% offer a comprehensive range of subjects.

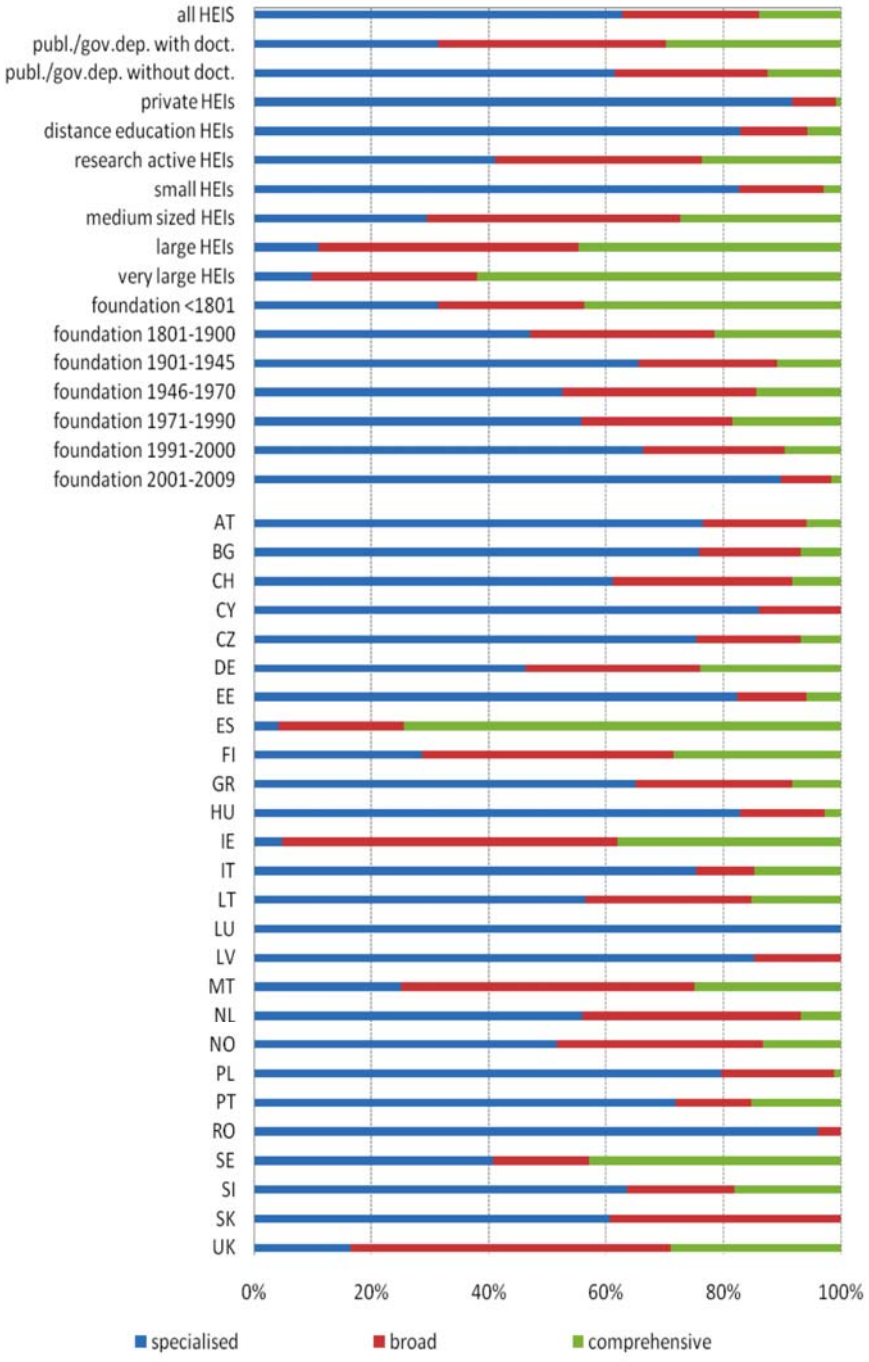


Figure 19. Range of subjects of the European higher education institutions by category and country

Source: EUMIDA. Dataset does not include Denmark and France (no data available).

Subject fields covered by the majority of European HEIs are "Social sciences, Business and Law" and "Humanities and Arts"; the fields of education "Science", "Engineering,

manufacturing and construction", "Education", "Health and welfare" and "Services" are covered by approximately one third of HEIs, while subject fields "General" and "Agriculture" are least common, being covered by 15% and respectively 11% of HEIs.

Historically, the process of establishment of new HEIs reflects the emergence of changing societal demands. Considering the range of subjects, the majority of the still existing very old HEIs (founded before 1801) offer qualifications in a comprehensive range of subjects. On the contrary, HEIs founded in the periods afterwards, are predominantly specialized. For example, technical universities were mainly founded in the 19th century or later, while a lot of the youngest, specialised HEIs focus on "Social sciences, Business and Law".

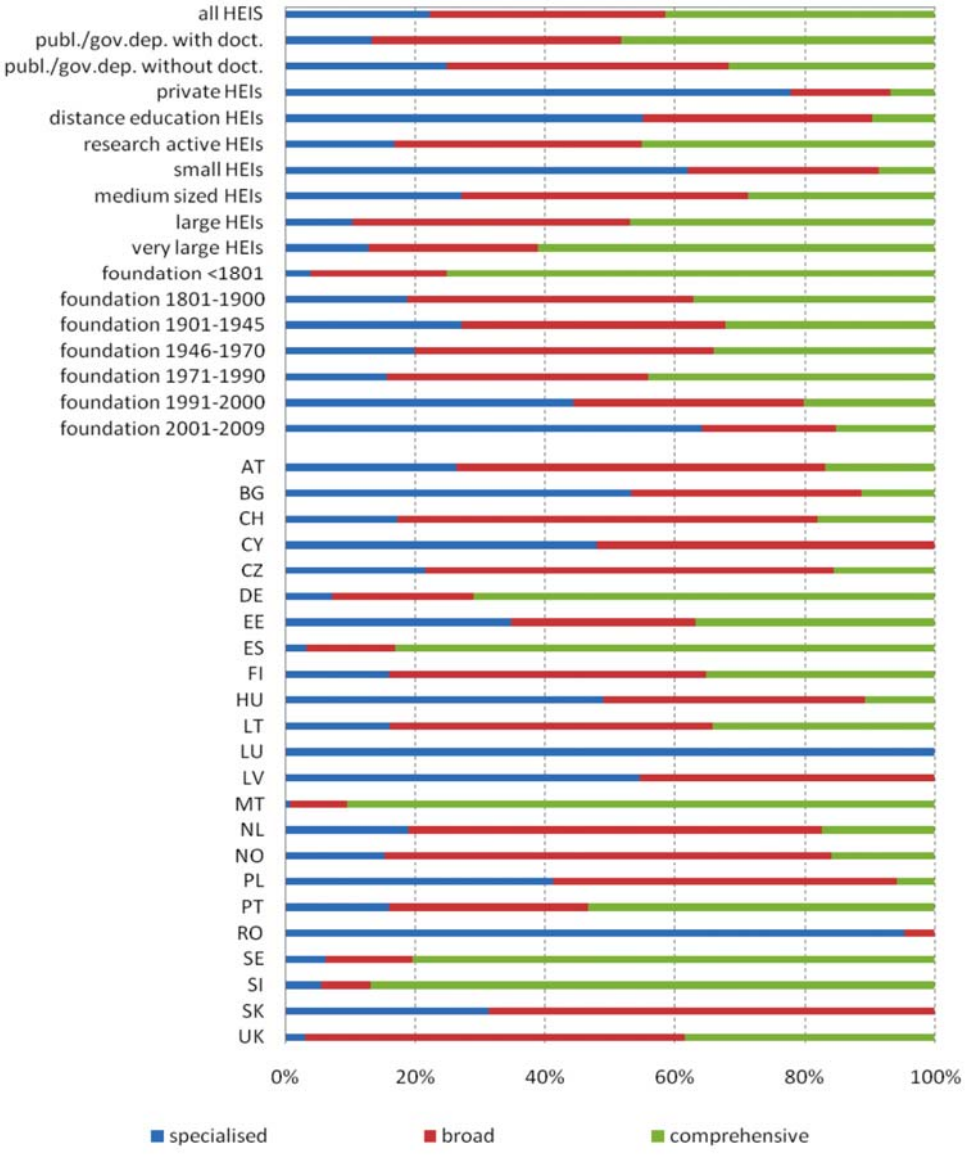


Figure 20. Range of subjects of the European higher education institutions (weighted by student body) by category and country

Source: EUMIDA. Dataset does not include Denmark and France (no data available); Belgium, Greece, Ireland and Italy (limited data availability).

A more balanced picture emerges when the distribution of institutions is weighted by the number of students: only 22% of students are enrolled in specialized HEIs, while 37% and 41% study in institutions with a broad and comprehensive subject mix, respectively.
However the majority of students in small, very young or private HEIs is enrolled in specialized institutions, as expected.
Countries that are dominated by specialized HEIs in terms of the student body are the Eastern European countries Bulgaria, Latvia and Romania. The University of Luxembourg is also specialized, offering degrees in the three fields of "Science", "Humanities and Arts" and "Social sciences, Business and Law". On the other hand the higher education systems of Germany, Portugal, Sweden and Slovenia are dominated by HEIs offering a comprehensive range of subjects. This suggests that EU countries followed very different strategies in developing their higher education systems.

6.6 International orientation

On average the share of international students at level ISCED 5 is 7%, while the median is significantly lower at 2.5%. Slightly less than 15% of institutions have no international students at level ISCED 5. All of these are small or medium-sized. On the other hand, the group at the other extreme, with a share above 25%, does not seem to be contingent on the type of HEIs. However, the group of old HEIs (founded in 1900 or before) is dominated by institutions with a share of international students above 5%.
The share of international students at level ISCED 5 is larger in some smaller countries (Austria, Belgium, Netherlands, Sweden and Switzerland), in Germany and United Kingdom, and in Cyprus and Luxembourg, the latter two representing special cases. It is very low in Eastern European countries (Estonia, Latvia, Lithuania, Poland, Romania, Slovenia), as well as in Spain.
The reader is referred to Sections 7.2 through 7.4 of this Report for further evidence on the patterns of internationalisation.

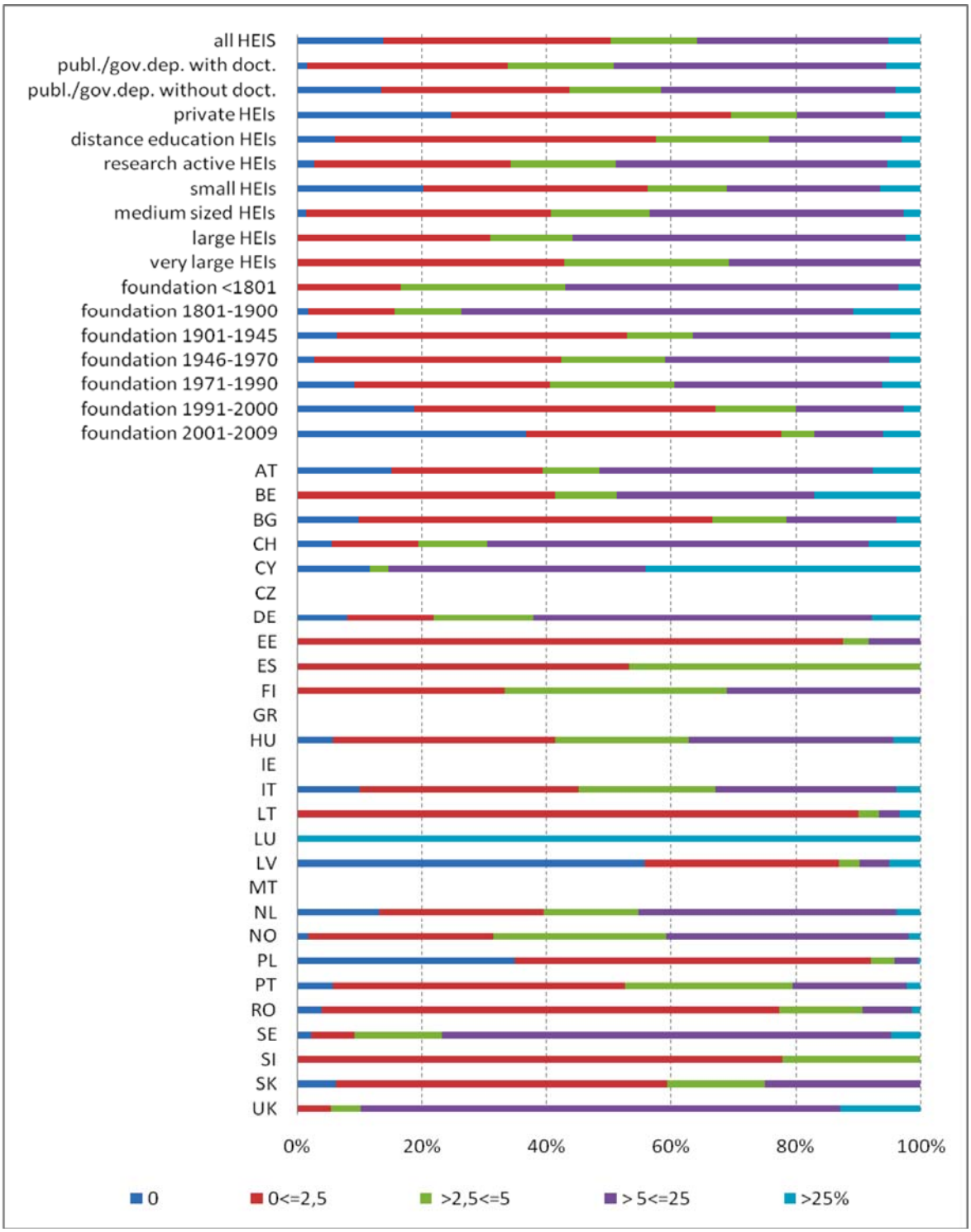


Figure 21. Share of international students ISCED 5 of the European higher education institutions by category and country

Source: EUMIDA. Dataset does not include Denmark and France (no data available), Czech Republic, Greece, Ireland, Malta (limited data availability).

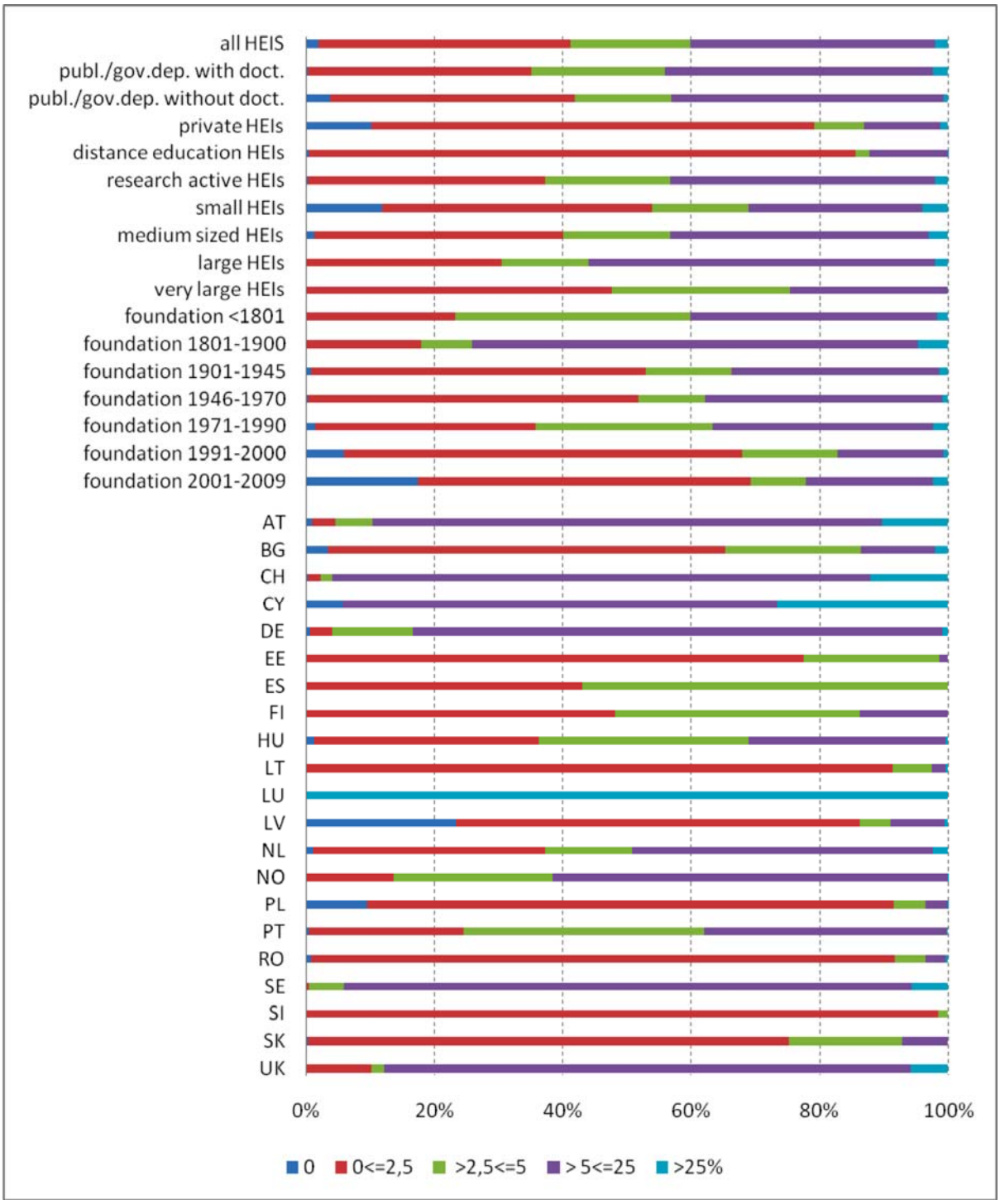


Figure 22. Share of international students ISCED 5 of the European higher education institutions (weighted by student body) by category and country

Source: EUMIDA. Dataset does not include Denmark and France (no data available).

6.7 Research activity

As already noted, there is not obvious overlapping between institutions granting the doctorate degree and those that are research active, the latter class being much larger. Among research active institutions, 39% are not awarding doctorate degrees. It is again useful to disentangle this phenomenon by country, since there are large national differences here.

First, there are several countries in which all institutions (or more than 95%) are considered research active. These are Austria, Belgium, Finland, Ireland, Luxembourg, Netherlands, Spain, Switzerland, and United Kingdom. Among them we find two opposite cases: countries that have a robust non-doctoral sector, and countries such as Ireland, Spain and United Kingdom that have not included, for different reasons, non doctorate awarding institutions in their higher education sector.

Second, a robust non-doctoral sector may mean that there are countries in which around two thirds of research active institutions are represented by institutions that do not grant the doctorate. These are again Austria, Belgium, Netherlands and Switzerland. In Germany and Hungary this share exceeds 60%.

This is an important finding, because it sheds light on the size of the research-performing, non-university sector in Europe.

The existence of a non-doctoral university research sector seems to be a consequence of the institutionalization of a relatively strong sector of education outside the traditional university model (granting doctorate degrees), which over time developed research competencies and demanded institutional legitimisation in the research field. The problem of the mission and capabilities of such a sector is a relatively new one, which deserves further investigation.

6.8 Institutional labelling

It is clear from the previous discussion that the higher education landscape in Europe is populated by institutions that greatly differ by educational mission (highest degree), field of education (specialised, broad, comprehensive), as well as size of students and staff. This diversity is somewhat reflected into various denominations or labels.

The EUMIDA census collected the original names of institutions in the national language, and asked the National Statistical Authority to inform about an official translation into English. The translation may come from the NSA, or other official government sources. There is no official nomenclature at European level.

The following Section is a qualitative and preliminary contribution to the establishment of an official nomenclature that fully reflects the diversity of institutions. We started a preliminary classification using the various labels and crossing them by the degree granted by the institution. The following classification combines several criteria.

It turns out the following. First, there are several denominations that point to specialized institutions. These are most found in the fields of:

- Military and internal security
- Art and drama
- Economics and business
- Medicine
- Technology
- Theology
- Agriculture
- Teacher education
- Language
- Public Administration.

Overall, 709 institutions bring a name referring to a specialization within a single field. Among them, considering 4 missing data, 262 offer only bachelor degrees, 302 up to master degrees, and 141 are institutions granting the doctorate degree. Thus this group is heterogeneous with respect to the highest degree delivered. However, the group of specialized institutions is concentrated in the non-doctorate type.

Among institutions offering the doctorate, those that are active in just one field are 112. It is likely that among those with a specialized label there are some that are active in more than one field. Conversely, it is possible that truly specialised institutions do not bring an easily identifiable name, and/or are active in more than one field. To make an example, Università Commerciale Bocconi, a private university offering up to doctorate degrees in Economics, Business and Law, is not registered in the Economics and Business section below. Or it is likely that some technical universities or polytechnics (in the engineering sense) are recorded under the general category of universities. Future work might offer a full-scale tabulation of labels, degrees and fields.

The proportion between non-doctorate and doctorate institutions in these fields exhibits some differences across fields, with Art and drama, Economics and business, Theology, Language, Teacher education and Public administration oriented towards non-doctorate institutions, while Medicine, Agriculture and to some extent Technology are more balanced.

Second, there are 819 non-doctorate institutions that do not bring in their denomination an explicit reference to specific education fields. They are collectively labelled "College model": 532 deliver the bachelor degree, 236 the master degree. A few of them (n= 8) are labelled together non-doctorate institutions in the miscellaneous group "Other higher education institutions" but happen to deliver also doctorate degrees. Other institutions appear to deliver doctorate degrees, for a total n=21. By inspecting these cases, it is clear that they have an extremely low number of doctorate students. With the exception of these 21 units, none of them deliver doctorate degrees. This category might be considered the one of generalist non-doctorate institutions, while the former category, discussed above, includes mostly specialized institutions.

In this category the largest group (n= 264) is formed by Universities of Applied Science, the international denomination for Fachhochschule or similar institutions. The second largest group is formed by Other higher education institutions (n= 246). While the number of non-doctorate institutions is remarkable, as already stated the share of students enrolled in these institutions is relatively small with respect to doctorate institutions. Interestingly, in this category the label of "college" is seldom used.

Overall, the largest group (n= 861) is formed by institutions that are labelled Universities, or a variation thereof. As a matter of fact, some of them actually deliver only bachelor degrees (n = 41), while 106 end up with master degrees. On the other hand, doctorate delivering institutions are 706. Summing up, the label of university is neither a necessary nor a sufficient condition to identify an institution offering up to doctorate degrees.

Third, there are very few institutions (n=20) whose denomination suggests they are specialised in postgraduate education.

Finally, only 12 institutions are specialised in distance education and online education.

Table 45. Institutional labeling of Higher Education Institutions (HEIs) in Europe

Institution category (English)	n.a.	Bachelor	Master	Doctorate	Total
(Missing)	8	3	22	3	36
Military and internal security	**0**	**9**	**4**	**5**	**18**
Academy of the Ministry of National Defence			2	3	5
Military Academy				1	1
Military university college		4	1		5
Police university college			1		1
Public Military Polytechnic		1			1
Public Military University		4			4
University of Defence				1	1
Art and drama	**1**	**28**	**167**	**35**	**231**
			17		17
Academy of fine art			20		20
Art College	1	21	17	13	52
Drama School		2			2
Fine Art Academy		2	1	18	21
Free Academy			24		24
Higher Institute for Musical and Choreographic Studies			76		76
Higher Institutes for Artistic Industries			4		4
National Academy of Drama			1		1
School of Dance		1			1
University College of Arts		2	7		9
University of Arts				4	4
Economics and business	**0**	**56**	**29**	**12**	**97**
Academy of Economics		44	29	10	83
Advanced School of Tourism Education		2			2
Merchant Marine Academy		10		2	12
Medicine	**0**	**2**	**0**	**9**	**11**
Advanced Nursing School		1			1
Medical Academy		1		8	9

Medical University				1	1
Technology		**74**	**56**	**43**	**173**
Federal Institute of Technology				2	2
Institute of Technology			1	13	14
Polytechnic			28		28
Private Polytechnic		53	3		56
Public Polytechnic		16	4		20
Technical University		5	4	28	37
Technological Educational Institution			16		16
Theology	**1**	**27**	**16**	**15**	**59**
College (church-owned)		12	9		21
Higher Ecclesiastic Academy		4			4
Institution				1	1
Theological Academy		10	1	4	15
Theological College	1	1	5	8	15
Theological University			1	2	3
Agriculture	**0**	**0**	**1**	**7**	**8**
Agricultural Academy			**1**	**7**	**8**
Teacher education	**1**	**8**	**28**	**13**	**50**
	1				1
Teacher Education School		2	9	7	18
University college of teacher education			14		14
University of Education			1	5	6
University of further education				1	1
University of Teacher Training		6	4		10
Language	**0**	**24**	**0**	**2**	**26**
Higher School for Language Mediators		23			**23**
University for Foreigners		1		2	**3**
Public administration	**1**	**34**	**1**	**0**	**36**
University of Public Administration	1	33			**34**
Academy of the Ministry of Interior and Administration'		1	1		**2**
College model	**30**	**532**	**236**	**21**	**819**
	1				1
Branch of foreign higher education institution		1	1		2
College	1	47	1		49
Independent college		9	2		11
Institution of Tertiary Education	1	22	8		31
Juridical persons established college		8			8
Non-university type		28	15	2	45
Other Higher Education Institution		174	64	8	246
Other university institution		1	1		2
Private Institution				1	1
Private university college		3	2		5
Private university college with accredited courses		10	6		16
Professional higher education institution		16	5		21
Single Higher Education Institution	2	3	2		7
State college		18			18
State university college		1	22	3	26
University College		8	51	7	66
University of Applied Sciences	25	183	56		264
Graduate education	**0**		**1**	**19**	**20**
Academy of sciences				2	2

Higher School				6	6
Physical Academy				6	6
Postgraduate institution			1	1	2
Research institute				4	4
Online and Open university	**0**	**2**	**8**	**2**	**12**
Online University		2	8	1	**11**
Open University				1	**1**
University model	**8**	**41**	**106**	**706**	**861**
				7	7
Generalist University			3	13	16
HEI		3	4	1	1
Juridical persons established institution of higher education		1	8	5	14
Private University		22	20	12	54
Public university				36	36
Specialized higher school		1	6	9	16
Specialized university				5	5
Specialized university, private				3	3
State institution of higher education			4	9	13
University	8	14	59	580	661
University type			2	26	28
Total	**50**	**840**	**675**	**892**	**2457**
%	**2,0**	**34,2**	**27,5**	**36,3**	**100,0**

Source: EUMIDA. Our elaboration from DC1 dataset.

Table 45 is only a starting point to place preliminary order in the rich heterogeneity of labels use by HEIs in various countries. It should be subject to extensive field validation, to control for cases in which the denomination might not correspond (perhaps for historical reasons, or legal aspects) to the category introduced in the table.

6.9 Profiling the European higher education landscape: A cluster analysis

6.9.1 Does a European university model exist?

As discussed in Chapter 1, there is a large debate on the adequacy of the European university model with respect to the needs of knowledge society and economy. But does an entity like "the European university model" really exist? If it does exist, what are its features?

The question of (a) whether there is a common European HEI model and (b) how it looks like is empirically not easy to tackle, because it is actually a composition of two questions. The empirical problem set up by these two questions can be translated as follows: we are looking for something, of which we know neither whether it exists, nor how it looks like.

However, if we consider what a European HEI model would imply, we get a bit closer to developing a sensible empirical approach. In particular, the existence of a European university model would imply that the HEIs in Europe (despite for example certain differences) resemble each other in one or the other way.

This fact, however, can easily be translated into the wording of cluster analysis: if there were a European HEI model (over a given set of characteristics) we would expect that:

- The European HEIs can be grouped into a limited set of clusters
- These clusters do not differ from Member State to Member State.

In other words, the hypothesis that there is a European HEI model is equivalent to the hypotheses that there are no country-specific clusters.

Unfortunately traditional cluster analysis (e.g. standard hierarchical models) cannot be used to analyse the hypothesis of the absence of country-specific clusters, because the number of clusters must be exogenously provided by the analyst. If for example, the analyst asked for a one-cluster specification, then, by definition, there would be no country-specific clusters. Furthermore, since the validity of the cluster number cannot be tested, all results remain more or less arbitrary.

Therefore, if clustering was to provide a sensible means of testing the question of the existence of a European university model, it should be a method that does not only classify the observations into a given number of clusters but it should determine the number of clusters simultaneously (i.e. endogenously).

Only lately have such models – usually referred to as model-based clustering – become available (Dasgupta and Raftery 1998, Fraley and Raftery 1998, 1999 & 2002). These methods are based on the notion of finite mixture distributions, where a mixture distribution is a weighted sum of underlying distributions – the so-called components. Although mixture distributions are commonly used for density estimation, they can also be applied to cluster analysis. In particular each of these components can be regarded as a separate cluster,

because observations that are drawn from the same component share the same underlying data generating process (the statistical definition of a cluster).
The most prominent advantage of looking at cluster analysis in this way is that this method receives a statistical meaning implying that different models can be compared and their adequateness can be tested. In particular, it is possible to check what the optimal number of clusters is by using standard model selection criteria (e.g. the Schwartz-Bayes or the Akaike-Information Criterion). The suggested models are implemented in the mclust-package for R statistical environment, which is described in Fraley and Raftery (2007).
Given the cluster configuration, it is easy to check whether country-specific clusters show up or not, simply by analysing the distribution of national affiliations in each cluster. A bit more advanced than to use simple counts is to rely on specialisation ratios, which also take into account that each country has different weight in the overall sample.
More specifically a specialisation ratio for e.g. Germany compares the share of German institutes in the specific cluster to its overall share in the sample. If the former is larger than the latter, we say that Germany would be overspecialised in the sample. Based on a specific transformation used, the specialisation ratios are between 0 and 1, if a country is overspecialised, between -1 and 0, if it is underspecialised, and zero, if it is neither over- nor underspecialised. Since recently (Schubert and Grupp 2009) it is also possible to perform asymptotically valid inference for these indicators given the observations are independent. Where applicable, this has also been performed, which for example allows checking whether a given specialisation is statistically significantly different from zero or not.

6.9.2 Measuring the dimensions of HEIs

In order to detect similarities and dissimilarities in activities using the proposed cluster approach, we have to define the relevant dimensions of the behaviour of HEIs. We do so in terms of a production-logic; i.e. we ask for the outputs and some characteristics of the HEIs. In an ideal scenario we would therefore include indicators for their three missions; that is teaching, research and knowledge generation as well as technology transfer (Schmoch and Schubert 2009b, Schmoch *et al.* 2010; compare also Gulbransen and Slipersaeter 2007).
In our case, we do not have variables for knowledge and technology transfer. However, we might hope that the legal status might be a rough proxy, because private HEIs are very often much more tied to private business. With respect to research-activities, we do not have available a publication-related measure, but there is a classification which indicates whether a university is research-active or not.
Other important characteristics certainly include size, internationalisation, degree of specialisation in terms of subject and legal status.

In summary, we use variables depicting legal status (private yes/no), the size (number of ISCED 5 and ISCED 6 students)[25], the teaching intensity (number of ISCED 5 students per staff), the graduate-teaching intensity (number of ISCED 6 students per staff), the internationalisation (share of international ISCED 5 and ISCED 6 students), the number of subjects covered (simple count over 9 distinct subjects), and an indicator for research-activity (Research-intensive yes/no).

This is a modelling exercise mostly oriented towards teaching activities of HEIs. Only the dummy variable for research activity is included. We present below the results of another exercise, in which two other variables are added to represent research activities.

6.9.3 Results

Based on the set of variables presented above the optimal Gaussian mixture distribution is calculated to be an ellipsoidal model with equal shape and two components.[26] Thus, a major finding is that there are only two clusters or different types of HEIs:[27]

Table 46. Characterization of two clusters of European HEIs

	University model	College model
Share of public institutions (%)	98,0	24,0
Number of students (ISCED 5 and 6)	12.541	2.054
Number of Students ISCED 5 per staff (ratio)	10,85	15,67
Share of international students (ISCED 5 and 6) (%)	9,0	4,0
Number of Students ISCED 6 per staff (ratio)	0,28	0,03
Number of subjects covered	5,12	2,14
Share of research iactive nstitutions (%)	99,0	15,0
Number of institutions in cluster	974	893
%	52,2	47,8

Source: EUMIDA.

[25] This is taken in logs because of excessively large numbers that could easily dominate the whole classification.
[26] The terms ellipsoidal and equal shape basically refer to features of the covariance between the three elements (components) of the mixture distribution. In our case, we find that the components are not independent from each other.
[27] We should note that, taking into account the restrictions made on the dataset and further losses due to item non-response, the cluster configuration still contains 1867 HEIs. This corresponds to roughly 75% of all HEIs in Europe. Therefore, all that follows should still give a representative picture of the European HEI landscape.

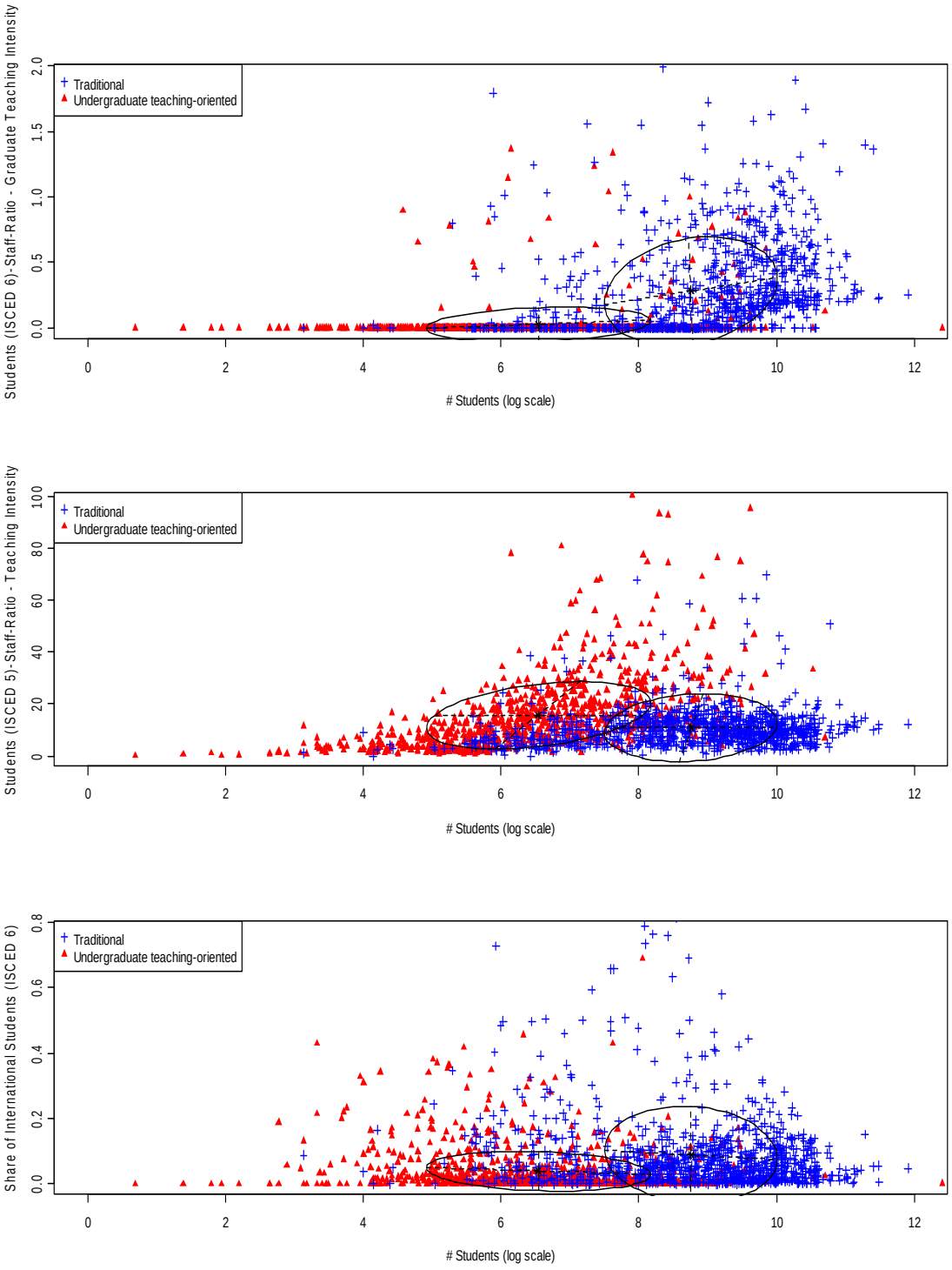


Figure 23. Clusters of Higher Education Institutions by Graduate teaching intensity, Undergraduate teaching intensity, Share of international Phd students.

Source: EUMIDA.

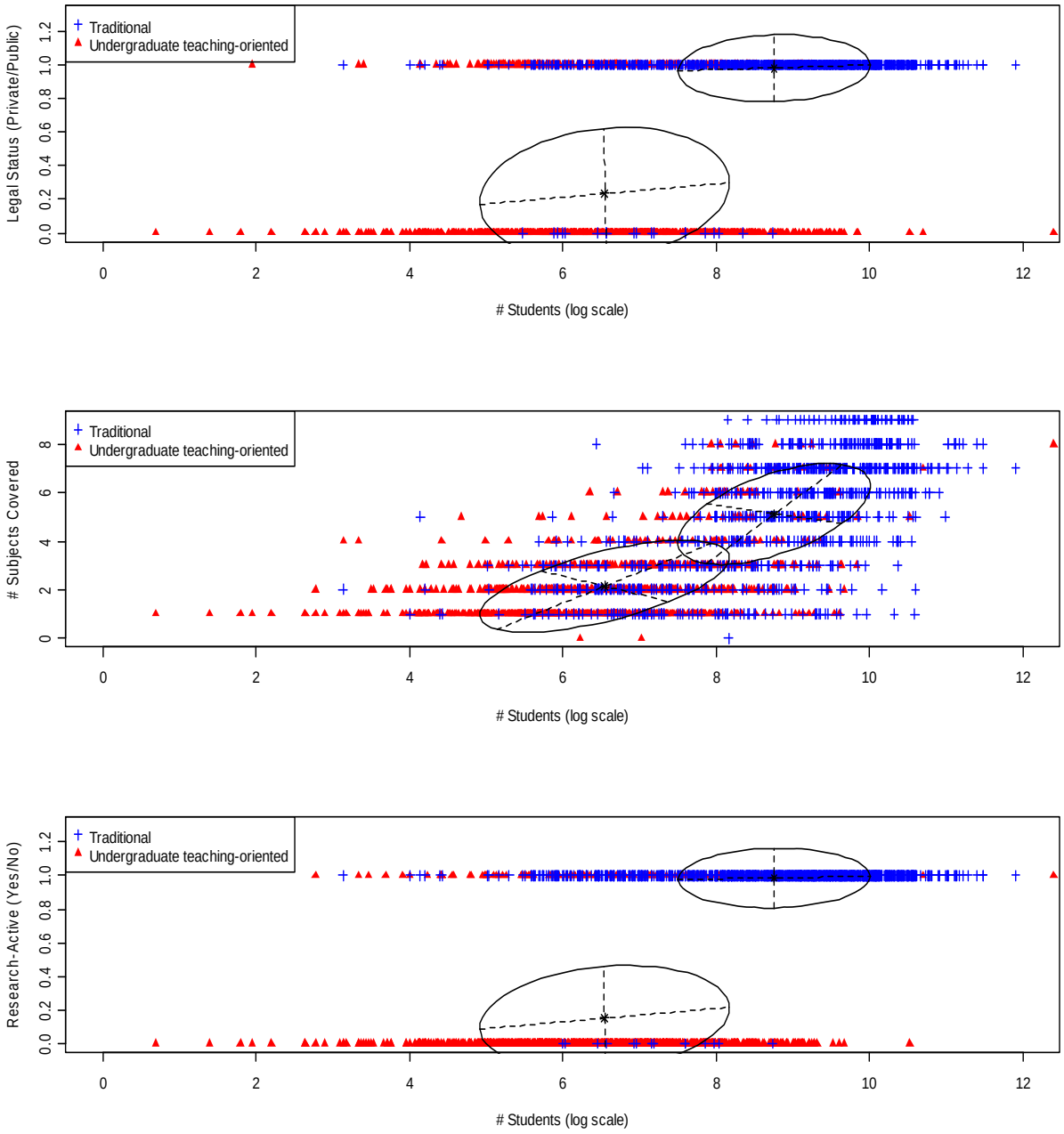


Figure 24. Clusters of Higher Education Institutions by Legal status (public/private), Number of fields covered, Research activity (yes/no)

Source: EUMIDA.

We tentatively call the two clusters "University model" (indicating that this cluster follows usual university activity profiles in an Humboldtian sense of unity of teaching and research) and "College model" (reflecting a cluster of HEIs that mainly focus on teaching, leaving aside research). The reasons for this wording can be read off Table 45, which includes the cluster means. For the sake of completeness, the graphical representation of the cluster configuration can be found in Figure 23 and Figure 24.[28]

In particular, we observe that the "traditional" cluster consists of public institutions only and contains on average the largest units (12,541 ISCED 5 and ISCED 6 students). It also has the lowest teaching load (10.85 ISCED 5 students per staff) and has an average degree of internationalisation in terms of students (9%). Furthermore, 99% of the members of this cluster are classified as research-active. On average they cover 5.12 different subjects.

Contrary to that stands the cluster of undergraduate teaching-oriented HEIs. They are with 2045 students much smaller. Furthermore, they hardly have any ISCED 6 students (0.03 per staff-member) and the majority of them is organised privately (only 24% are public). Additionally there are roughly 8% not active in research and rather focussed on terms of covered subjects (2.14).

These two clusters are in fact self-explaining, as they reflect traditional and polytechnical universities (universities of applied sciences); one of which is based on then Humboldtian ideal of the entity of research and teaching while the other is often focussed on few subjects and does not engage in research to the same extent.

6.9.4 National systems of higher education

Up to now we were able to show that the number of clusters necessary to sketch a map of the European universities is with two rather limited. Thus the European university landscape is – at least in terms of actual behaviour – not the jungle it is often believed to be. Furthermore, if it additionally could be shown, that the countries scatter over all clusters, this would give an indication that also country-specificities are not too important and that indeed something like a European university model might exist, despite large heterogeneity in terms of governance structure.

This analysis should be based on the country-shares in each cluster. However, there are basically two sensible ways of calculating them. Firstly, we could focus on the pure number of institutions, which is certainly relevant but hides that the HEIs can be of considerably

[28] Each of the Figures 23 and 24 a-c is a simple scatter plot, where on the x-axis the size is plotted (as measured by the number of students). On the y-axis the remaining four variables are depicted iteratively. The three ellipses in each diagram characterise the confidence ellipse for each cluster, where the center of each gives the means in both the variable on the y and the x-axis. What is easy to see, is that the clusters are allowed to be overlapping in some directions. For example in Figures 23 b and c the green cluster seems to scatter both over the blue and the red cluster. However, each of the diagrams is partial in that it only reflects two out of five dimensions. Therefore, any point that seems to be overlapping with another cluster in one diagram does not need to be in another.

different size possibly which, could possibly distort the results. Therefore, we secondly look at the country focus by number of ISCED 5 students.

It shall be noted at this point that the analysis will only be done for countries with roughly complete data, because any systematically missing data can easily distort the results on the country profiles. In particular this means that we will not report the results for the countries mentioned above (Austria, Belgium, Bulgaria, Finland, Ireland, Spain, and Switzerland).

Before we focus on the relative frequencies by cluster, we will shortly have a look to the absolute figures in Table 47.

Table 47: Number of HEIs and ISCED 5 Students by Country

	Number of HEIS	Number of ISCED5 students	Inclusion in Profile Analysis
AT	32	250,140	no
BG	35	206,999	no
CH	35	175,217	no
CY	22	26,403	yes
DE	352	2,068,787	yes
ES	47	1,134,826	no
FI	46	277,350	no
HU	70	390,453	yes
IE	21	156,881	no
IT	200	2,009,993	yes
LT	21	135,163	yes
LU	1	4,662	yes
LV	18	48,418	yes
NL	40	399,415	yes
NO	61	220,900	yes
PL	452	1,907,466	yes
PT	136	367,395	yes
RO	51	783,664	yes
SE	43	388,651	yes
SI	9	95,478	yes
SK	31	201,418	yes
UK	144	2,241,901	yes
Sum	1867	13,491,580	

Source: EUMIDA.

This reveals some interesting aspects: in terms of number of institutions Poland has the largest science system, followed by Germany, UK, and Italy. However, when looking at the number of students Poland is only on third place (after Germany, UK, and Italy), which indicates, that the majority of institutions in Poland is rather small. On the other side of the list, we have Luxemburg and Slovenia with 1 and 9 institutions respectively. In terms of students however, both Cyprus (26,403) and Latvia (48,418) are smaller than Slovenia.

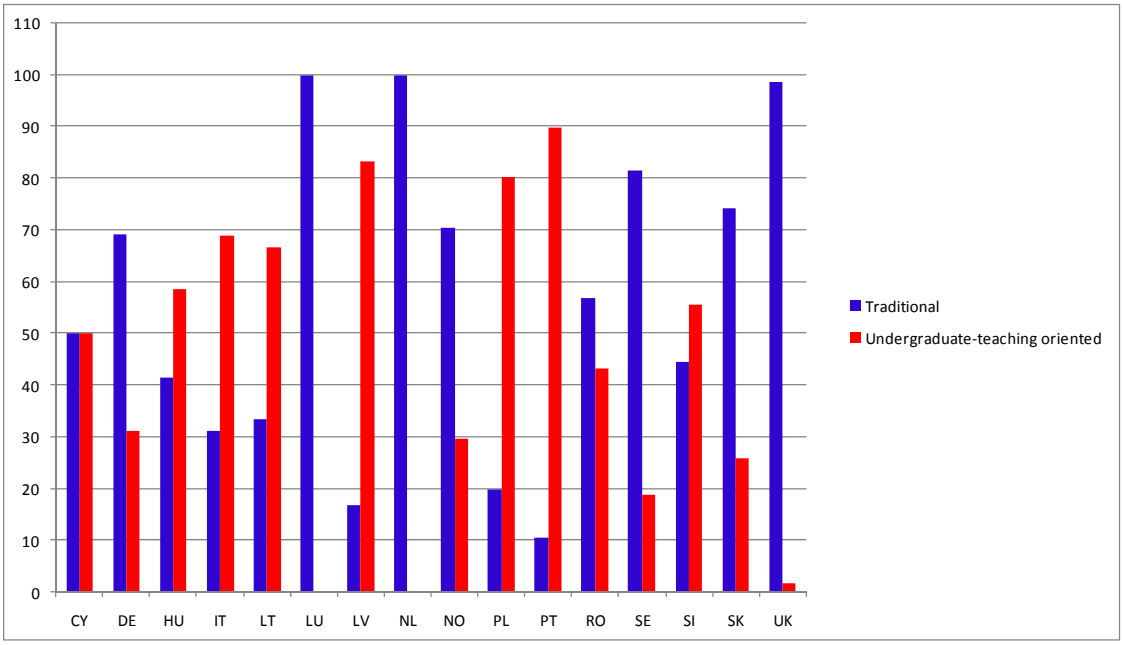


Figure 25. Share of Cluster by Country (Number of HEIs)
Source: EUMIDA.

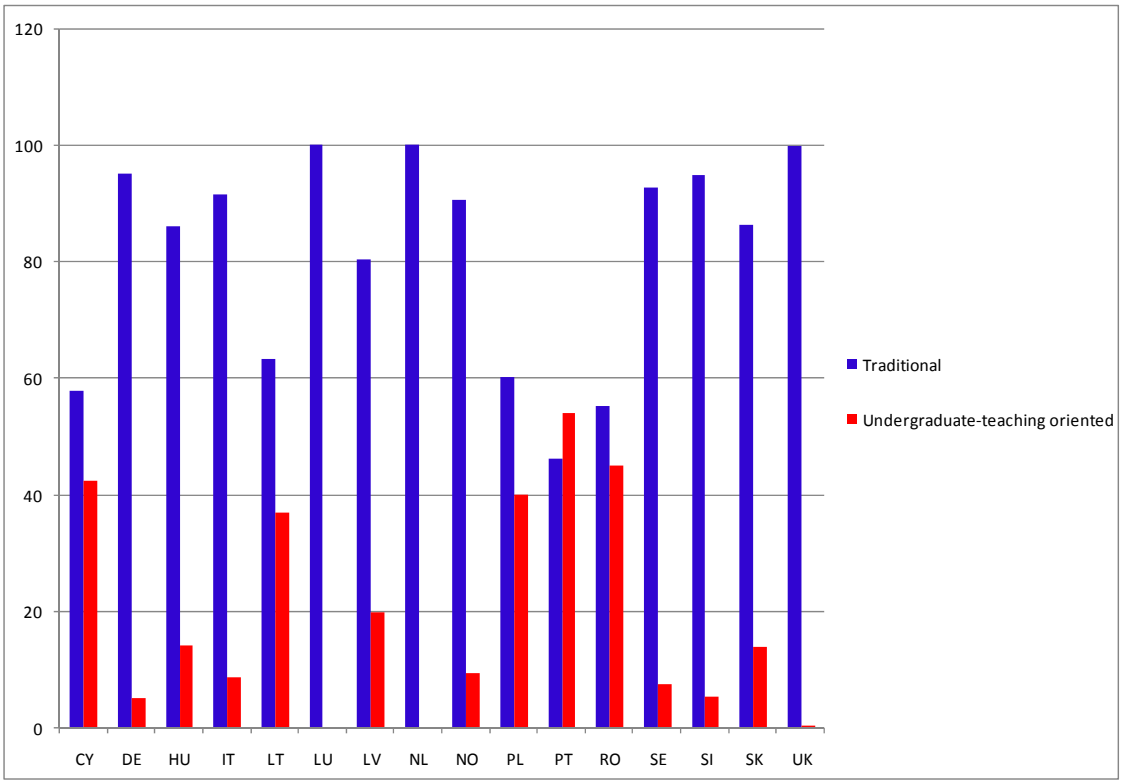


Figure 26. Share of Cluster by Country (Number of ISCED5 Students)
Source: EUMIDA.

Turning to the share of cluster by number HEIs (Figure 25) and by the share of students (Figure 26), we see that, although the number of HEIs that belong to either cluster is rather

levelled, the relative importance measured by student numbers lets the pendulum swing towards traditional HEIs. Except for the case of Portugal the majority of students are trained within the cluster of traditional universities.

Interestingly, we find some few countries that focus on the cluster of traditional universities exclusively or almost exclusively. Leaving aside Luxembourg (that has only one HEI in the sample), we see that the Netherlands and the UK show hardly any engagement in the cluster of graduate-teaching oriented HEIs.

On the other side of the spectrum we find many Eastern European countries (Latvia, Lithuania, Poland, Romania) and two Southern European countries (Cyprus and Portugal) that give much less weight to traditional universities.

In summary, this would lead us to say, that we find a relatively coherent image, since, first, we do not find any really country-specific, which would consist of only one or a few countries cluster (e.g. a cluster which ought to be labelled "Southern Europe"). Second, almost all countries have both traditional and graduate teaching HEIs that coexist.

In the next subsection we will enrich this analysis by portraying the country profiles, which visualise very intuitively the relative weights of the clusters given by each country.

6.9.5 Country profiles

The figures from above give a rough overview on several aspects of our questions. In particular we were able to show that country-specific clusters are absent, even though a few countries do not have HEIs in each of the two clusters. However, the profile of a country (i.e. the size-adjusted relative focus given to each cluster) can be more concisely described by specialisation ratios. As we will see, this highlights the fact, that although most countries are present in all clusters, they are with varying intensity.

As explained above a specialisation ratio measures the importance of a cluster for one country in terms of its weight in the overall sample. Thus, we call a country unspecialised, whenever its weight in the cluster is the same as in the overall sample. Using the tanhyp-Transformation no specialisation is indicated by a value of 0. A value of -1 corresponds to maximum underspecialisation, which occurs if none of the country's HEIs is member of this cluster. On the contrary a +1 implies maximum specialisation or means that all of the country's HEIs fall in this cluster.

Once again we calculate the specialisation ratios both based on the number of HEIs and on the total of ISCED 5 students. For the former case, we also report the significance-levels based on the mathematical theorems provided by Schubert and Grupp (2009).[29,30]

[29] Unfortunately, the theorems do not apply to the specialisation ratios based on student counts, basically, because the latter are clustered by university. The methods provided in Schubert and Grupp (2009), however only work for mutually independent data, which prohibits among other complications any form of clustered data.

[30] Note that the statistical inference is not available, whenever a cluster-specialisation of +1 or -1 occurs.

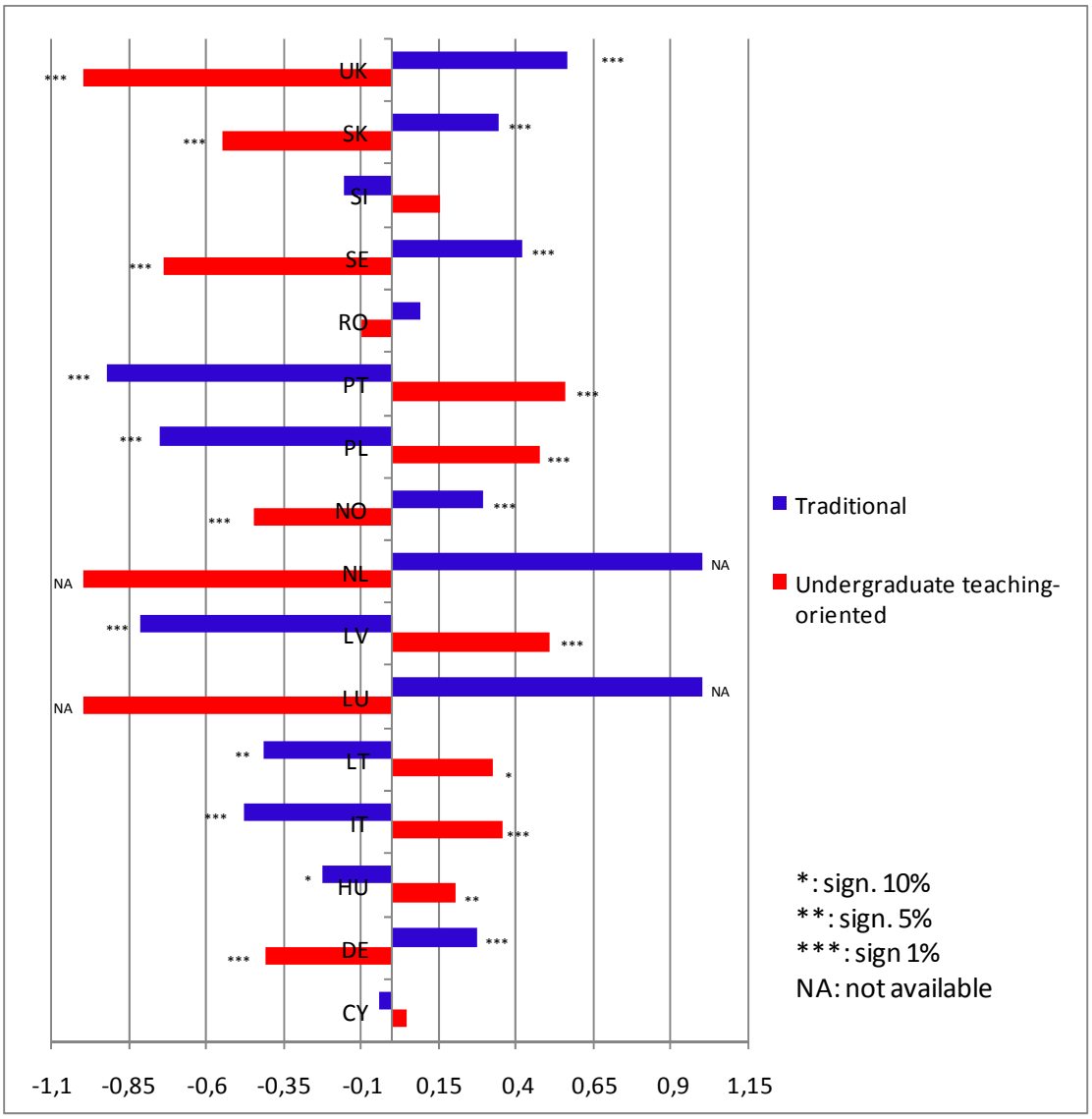


Figure 27: Country Profile (Specialisation Based on Number of HEIs)

Source: EUMIDA.

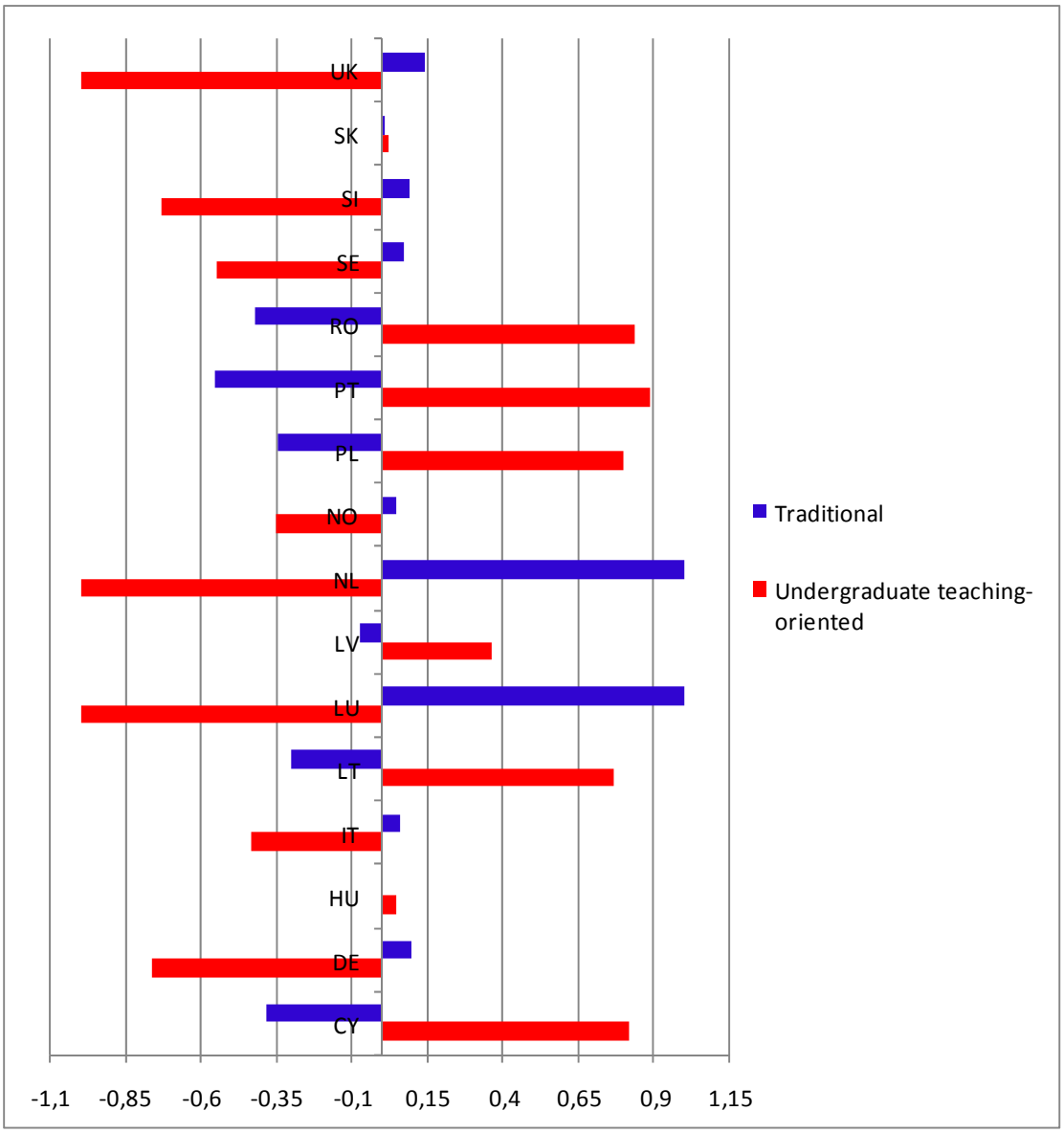


Figure 28: Country Profile (Specialisation Based on Number of ISCED 5 Students)
Source: EUMIDA.

As can be seen from Figures 27 and 28, there is – despite the results from the last subsection that highlight the importance of common features – also some heterogeneity in terms of the (relative) country profiles. As we already noticed in the last subsection (once again leaving aside Luxembourg with only one university in the sample), there are some countries that are heavily specialised in the traditional model, most notably the UK and the Netherlands. Somewhat surprisingly, although Germany is considered to be the homeland of the Humboldtian ideal of unity of research and teaching, it is only slightly (even though significantly; see Figure 28) specialised here.

On the other side of the scale, some countries put much emphasis on the HEIs that are characterised by undergraduate teaching. Apart from Cyprus, and Portugal, this is true

especially the Eastern European countries Hungary, Latvia, Lithuania, Poland and Romania. This seems to indicate that many Eastern European countries have focused on the teaching dimension of HEIs rather than traditional universities, characterised by the unity of teaching and research, which might be explained by the increasing demand for qualified personnel after the end of the Warsaw Pact (van Leeuwen and Foldvari 2008, compare also Mateju *et al.* 2003). In any case, it is not true for all Eastern European countries. In particular, Slovakia and Slovenia are underspecialised in the teaching-oriented cluster, at least with respect to students (Figure 28).

In summary, despite the results of the last subsection, which indicated the existence of HEI-types found all over Europe, we also find heterogeneity with respect to the relative importance of the clusters given by each country. In particular, we can find HEI systems that are more traditionally oriented (UK, Netherlands, and to some degree Germany) than others (Cyprus, Poland, Romania), while others have a roughly balanced mix (Hungary, Slovakia).

6.10 In search of the research university model

What happens if we add to the former clustering exercise other variables aimed at capturing the research dimension? In this Section we examine the results of a model in which we use the same set of variables depicted above and add the following: (a) ratio between number of ISCED 6 students and total number of students; (b) share of international ISCED 6 students.

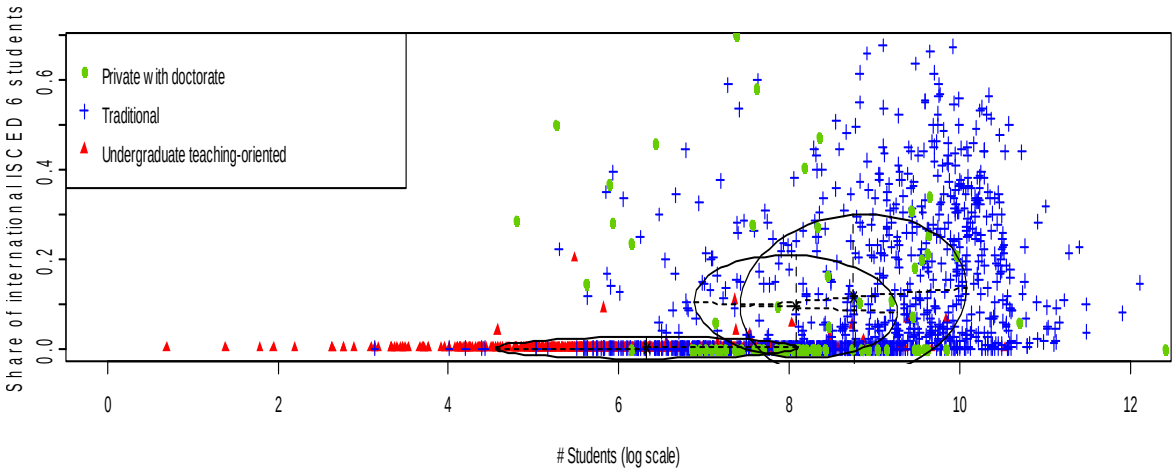
6.10.1 Results

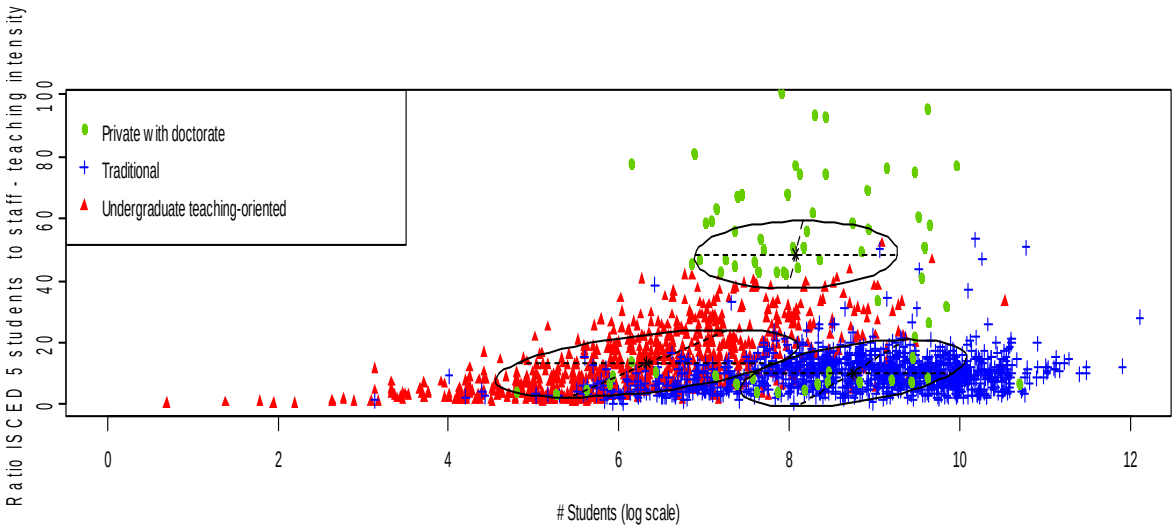
Interestingly, the structure of clusters is largely confirmed, while a third, small cluster emerges.

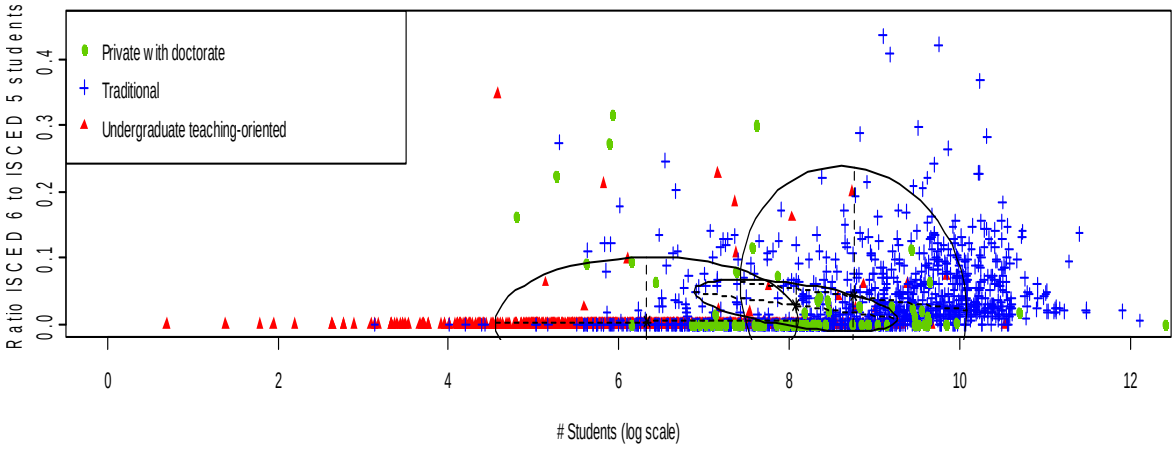
Table 48: Variable Means by Cluster

	Private with doctorate	Traditional	Undergraduate teaching-oriented
Share of public institutions in %	5.00	100.00	21.00
Number of ISCED 5 and ISCED 6 students	8792.82	12622.26	1330.65
Number of ISCED 5 students per staff	49.26	10.09	13.34
Ratio ISCED 6 to ISCED 5 students	0.03	0.04	0.00
Share of international ISCED 6 students in %	10.00	12.00	0.00
Number of subjects covered	2.37	4.62	1.80
Share of research-active institutions in %	54.00	96.00	13.00
n_i	76	800	992
n	1868		
Share of total sample in %	76.02		

Source: EUMIDA.


Private with doctorate
Traditional
Undergraduate teaching-oriented
Share of international ISCED 6 students
0.0
0.2
0.4
0.6
0
2
4
6
8
10
12
Students (log scale)


Private with doctorate
Traditional
Undergraduate teaching-oriented
Ratio ISCED 5 students to staff - teaching intensity
0
20
40
60
80
100
0
2
4
6
8
10
12
Students (log scale)


Private with doctorate
Traditional
Undergraduate teaching-oriented
Ratio ISCED 6 to ISCED 5 students
0.0
0.1
0.2
0.3
0.4
0
2
4
6
8
10
12
Students (log scale)

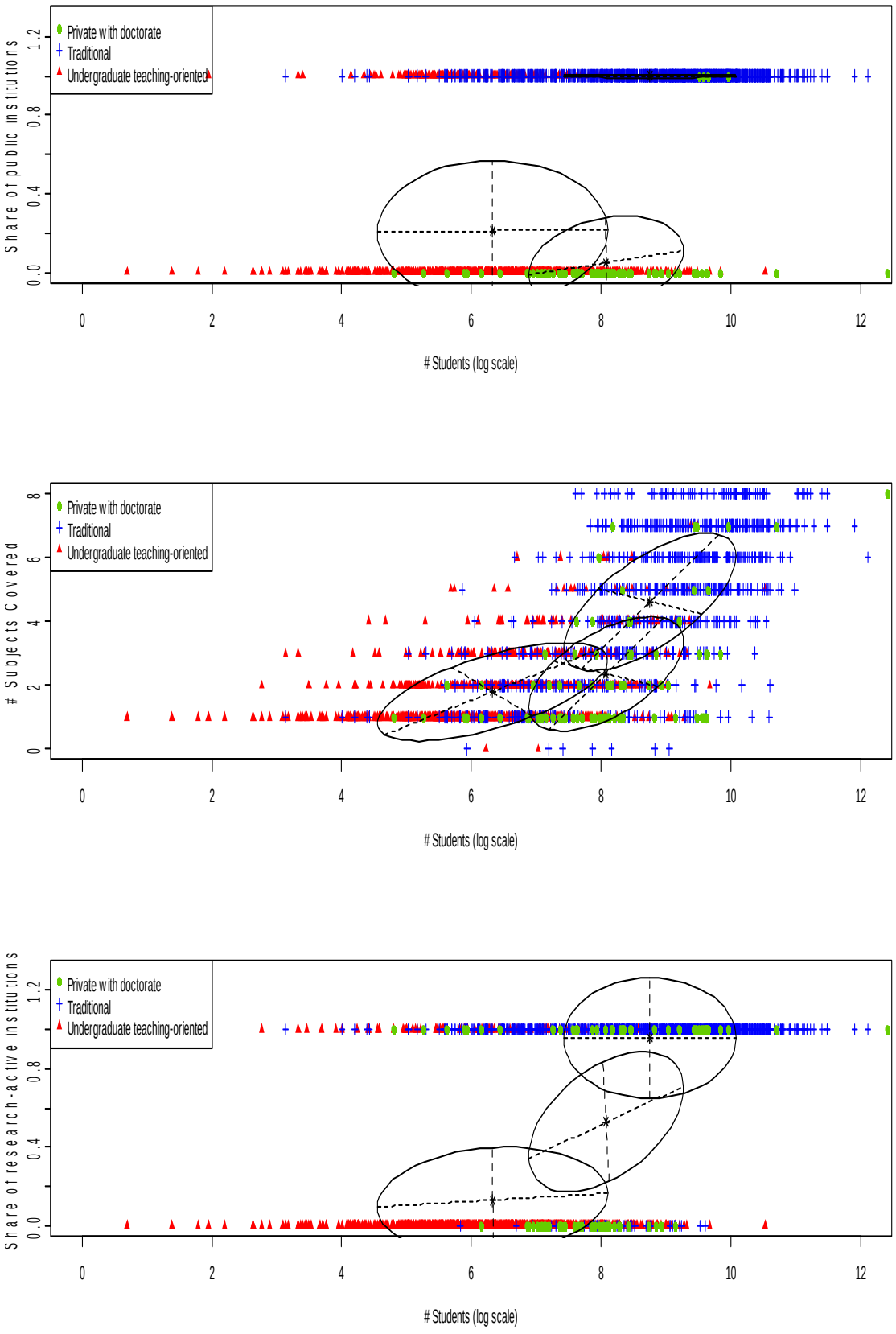


Figure 29a-f: Graphical Representation of the Cluster Configuration

Source: EUMIDA.

In particular, we observe that the cluster named "Traditional" consists of public institutions only and contains on average the largest units (12,622 ISCED 5 and ISCED 6 students on average). It also has the lowest teaching load (10.09 ISCED 5 students per staff). Furthermore, 96% of the members of this cluster are classified as research-active and they have the highest degree of internationalisation in terms of students (12%).[31] On average, the HEIs in this cluster cover 4.62 different subjects and are therefore rather broad. To summarise the results, this cluster conforms very well to a traditional (or Humboldtian) understanding of universities; e.g. relatively large institutes with a broad coverage of the disciplines, unity of research and teaching, and a strong science base.

Contrary to that stands the cluster of undergraduate teaching-oriented HEIs. They have with 1,330 students the smallest universities. Furthermore, they have no ISCED 6 students and the majority of them is organised privately (only 21% are public). Additionally there are roughly 13% not active in research and they are rather focussed on terms of covered subjects (1.8). The teaching load is slightly higher than in the traditional universities but still is with 13 students per staff of roughly the same magnitude.

The last cluster is a relatively small group of universities with seem in many respect to be a private copycat of the traditional universities. However, despite the fact, that their degree of internationalisation in PhD-training and their graduate-teaching-intensity are comparable to the cluster of traditional universities, they also have some distinctive characteristics. They are smaller in terms of students (8,797) and they are much more focussed in terms of subjects covered (2.37 on average). Furthermore not all of these institutions are characterised as research-active (only 54% are classified as such).

The traditional and the undergraduate-teaching oriented cluster are in fact self-explaining, as they reflect traditional and polytechnical universities (universities of applied sciences); one of which is based on then Humboldtian ideal of the entity of research and teaching while the other is often focussed on few subjects and does not engage in research to the same extent. The last cluster of private institutions is certainly less intuitive; and indeed looking at it more closely reveals that it reflects a large variety of private institutions that range from private business schools, church-run colleges, or private medical schools. Anyhow, this cluster is with 76 members corresponding to less than 5% of the sample of minor importance.

Apart, however, from what can be observed, it is also interesting to note what cannot be observed. In particular, the traditional cluster does not split up into "research" and, so to say, "regular" universities. This of course does not preclude that one university may be more research-intensive than another. But it means, in the parlance of cluster analysis, that the research-intensive universities are rather an extreme of the regular traditional universities rather than forming a cluster on their own. Based on this, it certainly makes sense to highlight differences in research-intensity between universities, but there is no support for

[31] This also corroborates the view of taking internationalisation as a measure of research attractiveness, as it was argued in the Report.

labelling one "research university" and another "regular university". In the end, all universities in the traditional cluster do in principle the same things, even if one may be more successful than another.

In summary, we find a three-cluster-configuration with traditional and undergraduate-teaching oriented universities (that account for 95% of all institutions) and a third very small cluster of private HEIs that also award PhD or doctoral titles.

6.10.2 Country focus by Cluster

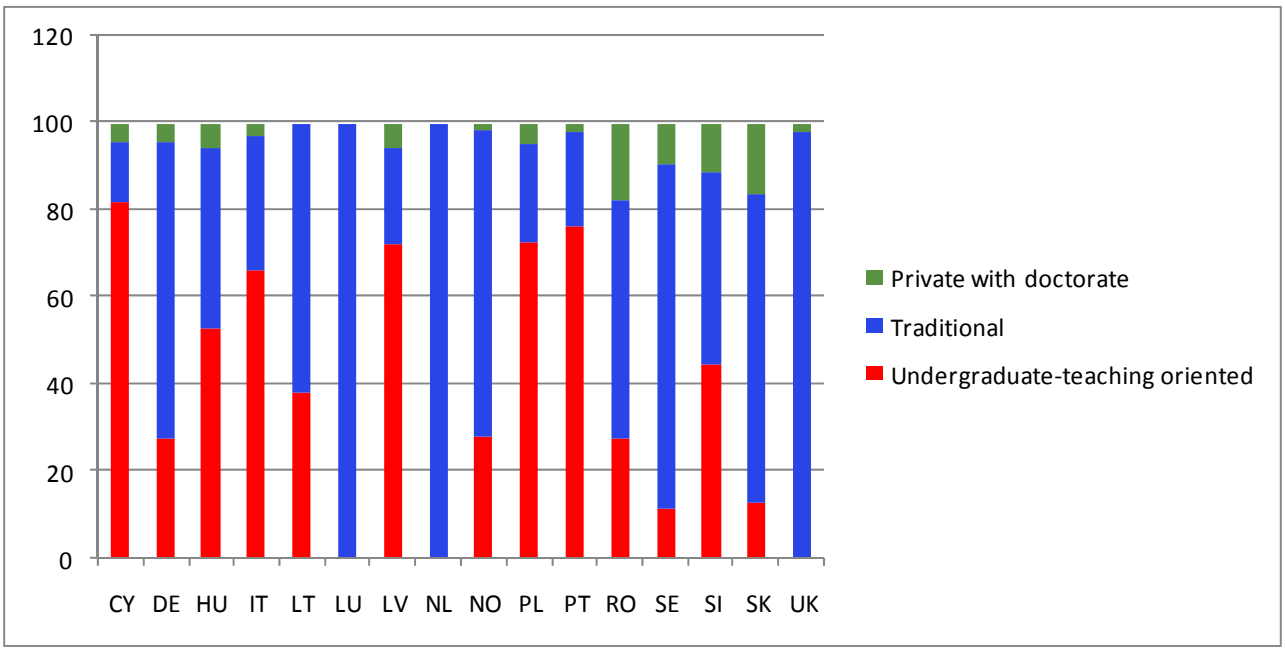


Figure 30: Share of Cluster by Country (Number of HEIs)

Source: EUMIDA.

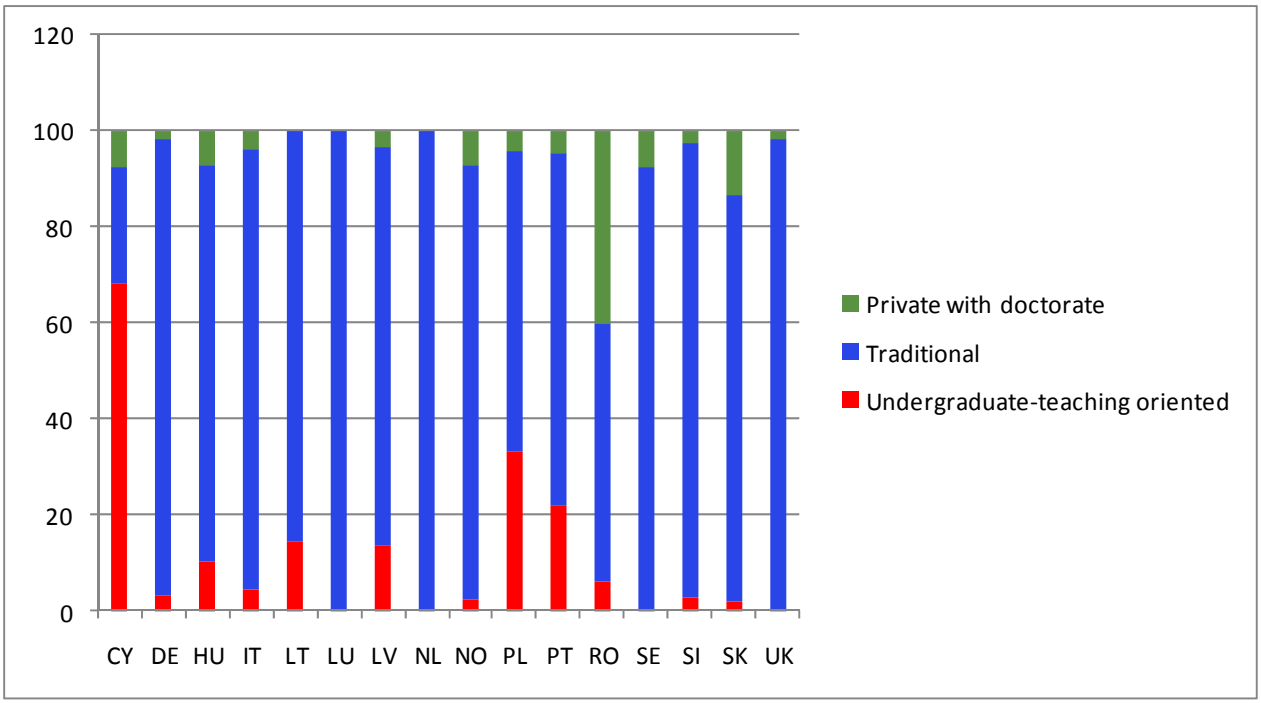


Figure 31: Share of Cluster by Country (Number of ISCED 5 Students)
Source: EUMIDA.

Turning to the country shares in the cluster by number of HEIs (Figure 30) and by the number of students (Figure 31), the picture discussed in the previous Section is largely confirmed.

The private institutions that award PhD or doctoral titles are rather marginal both in terms of numbers as well as in terms of students (except for the case of Romania where roughly 40% of the students are trained within this cluster of universities).

In summary, this would lead us to say, that we find a relatively coherent image, since we do not find a country or region-specific classification, which would replicate a geographic pattern. In the same vain, almost all countries have traditional, graduate-teaching-oriented and private HEIs that coexist.

In the next subsection we will enrich this analysis by portraying the country profiles, which visualise very intuitively the relative weights of the clusters given by each country.

6.10.3 Country Profiles

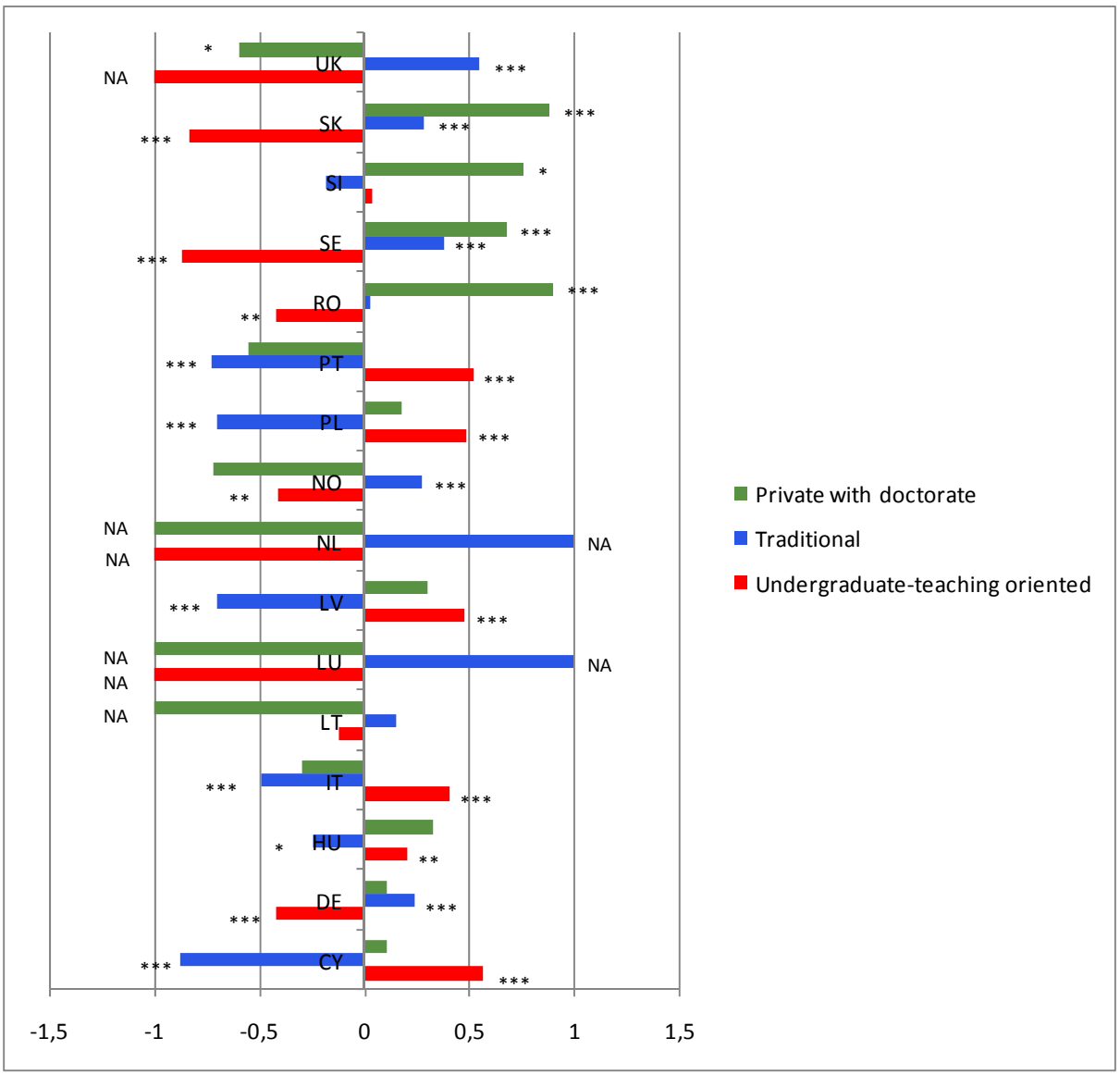


Figure 32. Country Profile (Specialisation Based on Number of HEIs)

Source: EUMIDA.

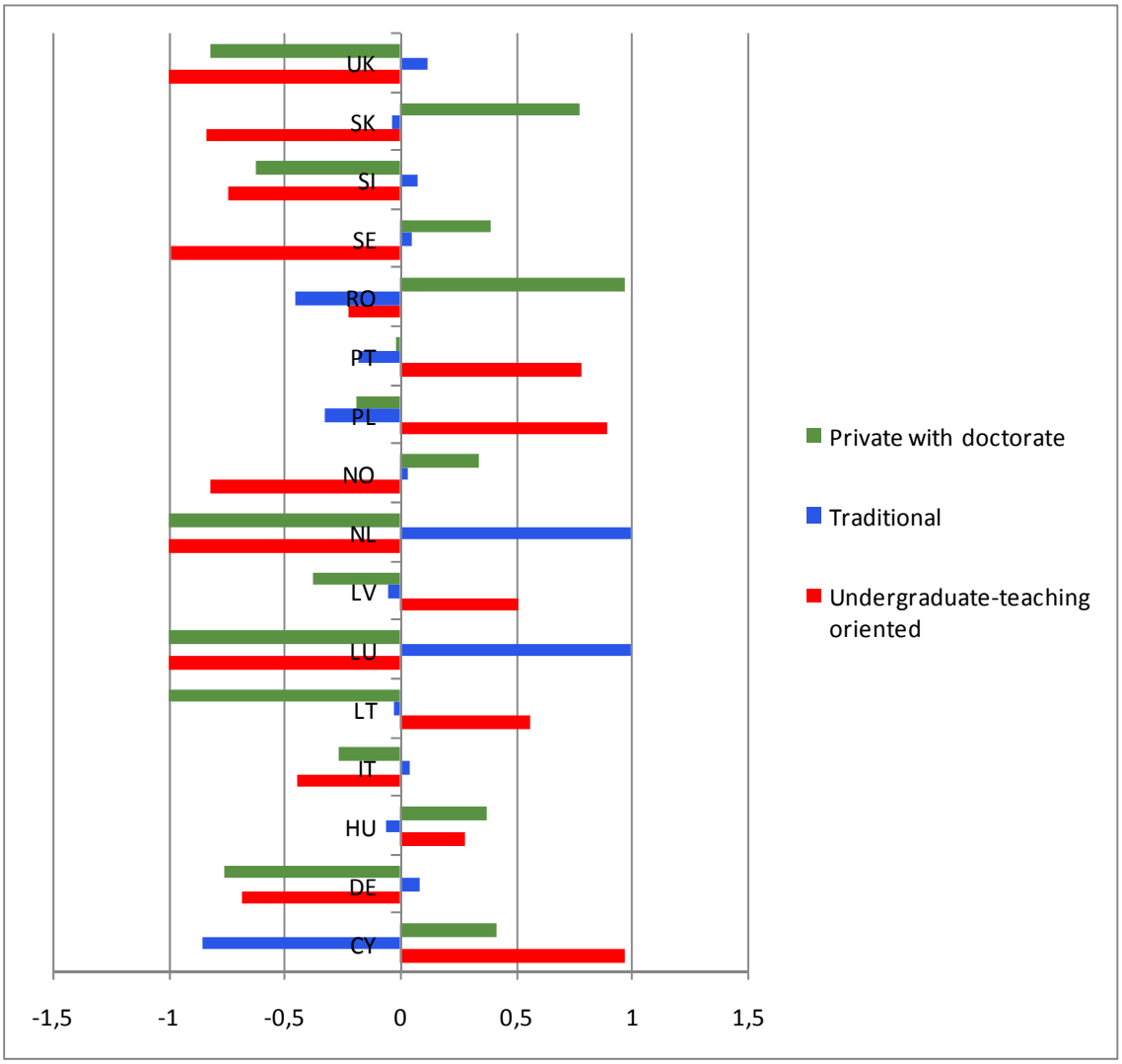


Figure 33. Country Profile (Specialisation Based on Number of ISCED 5 Students)
Source: EUMIDA.

As can be seen from Figures 32 and 33, country profiles are largely confirmed, with the new cluster of private specialised HEIs important only in Romania and Slovakia, somewhat less in Hungary, Norway and Sweden.

7. The structure of research-active higher education institutions in Europe

7.1 Identification of the research-active sector and status of Data Collection 2

As discussed at length in the previous Chapters, EUMIDA adopted a definition of research-active institutions that was based on the fulfilling of at least three criteria among a set of descriptions aimed at identifying those units with a permanent and legitimized organization for research. This choice has been made after excluding an approach based on thresholds - such as, for example, the intensity of PhD students.

As in Chapter 6, we present data for those countries for which all information is complete and reliable.

It must be underlined that the analysis of Data Collection 2 has been possible only after all information have been collected from National Statistical Authorities, which took place much later than planned. As a matter of fact, the first delivery of this Report took place on October 6th 2010, following the scheduling of the contract with the European Commission, while the EUMIDA Consortium research team received the last update from a few countries after September 20th, and received the data from Denmark at the end of September. Thus, the analysis contained in this Chapter must be considered preliminary: the EUMIDA team declares its willingness to support the Commission and EUROSTAT in the analysis of data on a voluntary basis well beyond the expiration of the contract.

In particular, in this Chapter it will not be possible to report on some variables included in Data Collection 2 - particularly, on Research expenditure, Total expenditure, and Revenues. This data require a careful analysis because they are subject to significant comparability problems (see Chapter 5), so that any country-level analysis would be premature and misleading. In this case the EUMIDA Consortium will continue to work on data.

7.2 Internationalisation of students

In the overall group of research-active institutions, roughly 50% have a share of undergraduate students coming from abroad of less than 5%. Internationalisation of students is greater in public universities, but the group of highly internationalized institutions (more than 15% of foreign students) is similar between public and private, between 15% and 20% of the total.

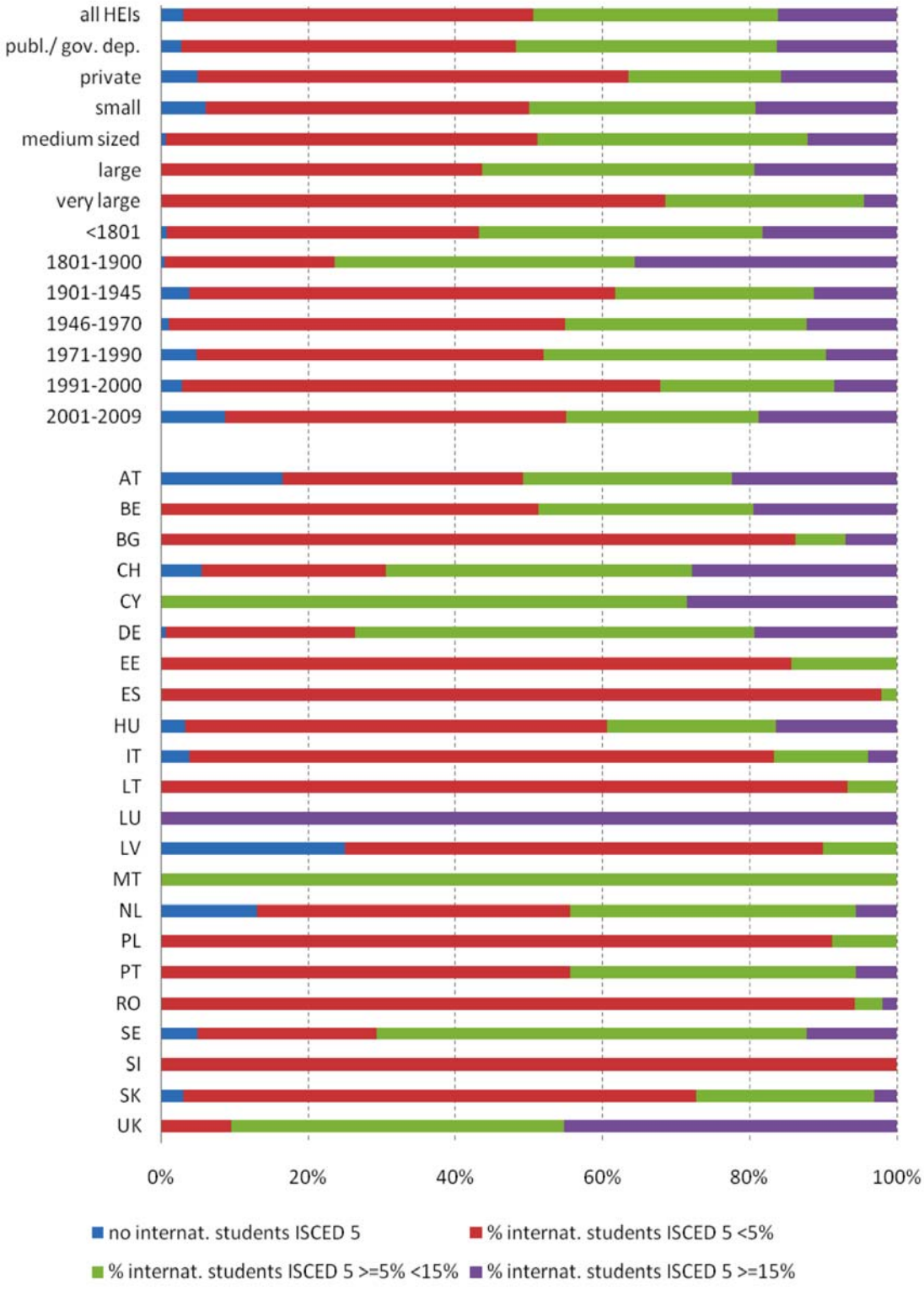


Figure 34. Share of international students ISCED 5 of research active European higher education institutions by category and country

Source: EUMIDA dataset 2010 excluding Denmark, France (no data available), Czech Republic, Greece, Finland, Ireland, Norway (limited data availability/ comparability)

With respect to size, the group of highly internationalized institutions is more important in small and large size classes, while is at the minimum in very large ones.
With respect to age, we find the most internationalized institutions among the very old (established before 1801), those created in the 19th century (almost 40% of the total in this period) and, interestingly, in the youngest ones, created after 2001. Thus it can be said that internationalization of students is polarized between very old and very young universities.
United Kingdom is the most open system, with more than 80% of institutions having more than 15% of foreign students.
Looking at cross-country differences, it is not surprising that, after UK, small countries have the highest degree of internationalization, as in the case of Austria, Belgium and Switzerland. However, Germany has also a share of highly internationalized institutions slightly above the average and a much larger share of intermediate level institutions (between 5% and 15%).
Among the large countries, Italy, Poland and Spain are the least internationalized.

7.3 Internationalization of doctorate students

As expected, the average level of internationalization is higher for ISCED 6 students than for ISCED 5, with around 30% of institutions having more than 15% of students from abroad and less than 20% having none.
Private institutions have a similar share of highly internationalized, but have a more than double share of zero internationalization institutions, with respect to the public sector.
Large institutions have a larger share of foreign PhD students, while very large ones are the least open, by a wide margin. The age pattern is similar to the one identified for undergraduate students: very old and very young institutions perform better. The former are most likely attracting PhD students due to their prestige and research track, the latter due to a proactive strategy to target the segment of mobile postgraduate students, whose size and mobility has greatly increased in the last few decades.
With respect to countries, United Kingdom and Switzerland stand out for a share of highly internationalized institutions in doctoral education exceeding 80%, a level far beyond other countries. Belgium, Austria, Sweden and Germany fall in the 30%-60% range of share of highly internationalized units. Eastern European countries are in general scarcely internationalized. Italy has a particularly poor performance, with around 5% of institutions having more than 15% PhD students from abroad.

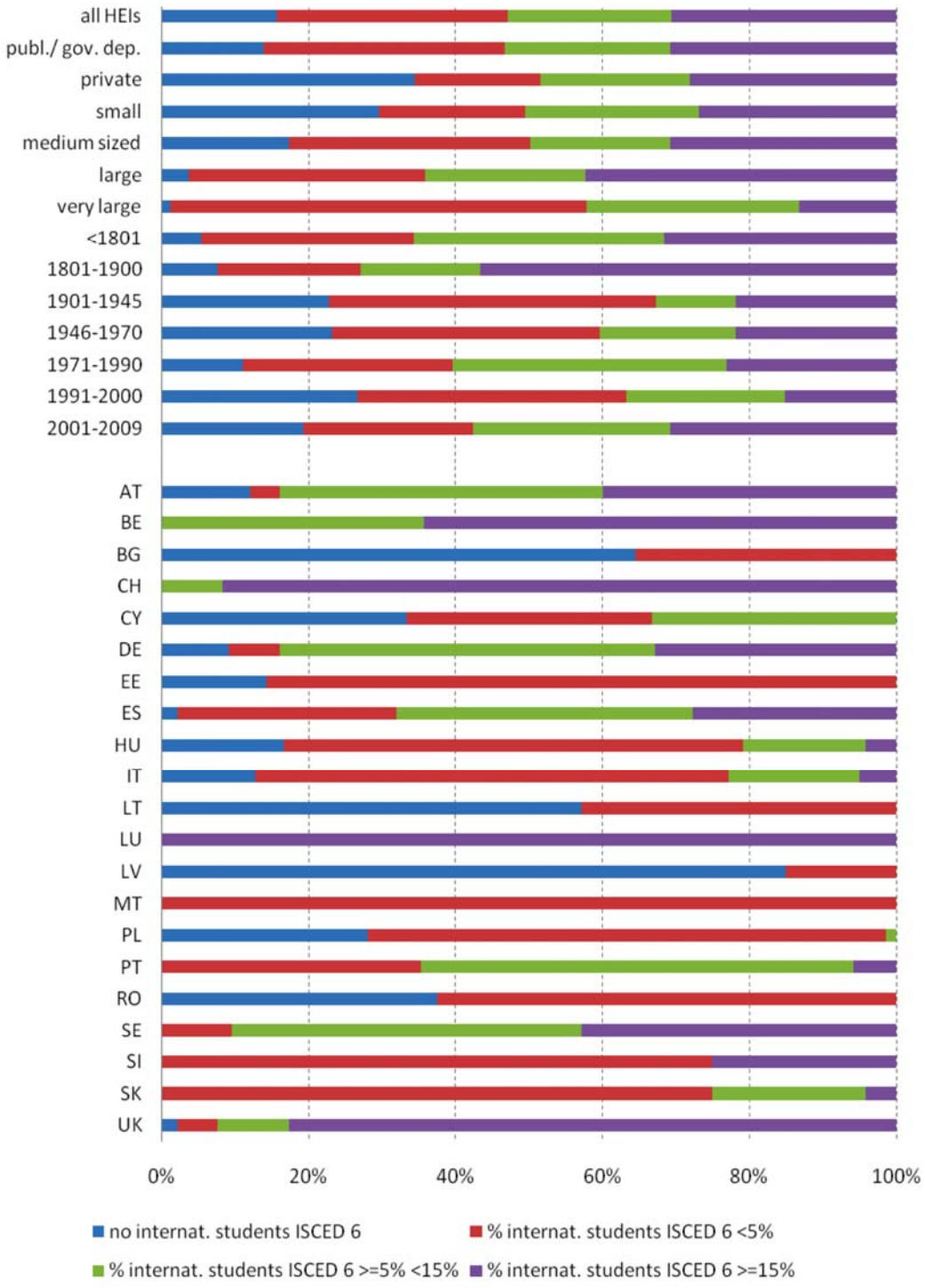


Figure 35. Share of international students ISCED 6 of research active European higher education institutions by category and country

Source: EUMIDA dataset 2010 excluding Denmark, France (no data available), Czech Republic, Finland, Greece, Ireland, Netherlands, Norway (limited data availability/ comparability)

7.4 Patterns of internationalisation

By combining data on international students in both ISCED 5 and ISCED 6 categories we obtain an interesting characterization.

We limit the analysis to the institutions (universities) awarding the doctorate degree and define the following categories:

- Broad international orientation: % of international students ISCED 5 & ISCED 6 > 15%
- International orientation ISCED 6: % of international students ISCED 5 <= 15%; ISCED 6 > 15%
- International orientation ISCED 5: % of international students ISCED 5 > 15%, ISCED 6 <= 15%
- National orientation: % of international students ISCED 5 & ISCED 6 <= 15%

Following this characterization, it turns out that more than 50% of European universities have a national orientation, while around 15% have a broad international orientation. In the middle, a few universities are internationalized only in undergraduate education, while 30% are open mainly in postgraduate education. Private, small and very large universities are more likely found in the national orientation category.

The combined ISCED 5 & ISCED 6 characterization is useful to examine the national differences. Broad international orientation is relevant in Austria, Switzerland, and United Kingdom, to a lesser extent also in Germany and Sweden.

Spain exhibits international orientation only for PhD education and no institutions with broad orientation. Italy has very few broadly open institutions, but a much lower average orientation. All Eastern European countries have a largely national orientation, with a limited exception for postgraduate education in Slovakia and Slovenia.

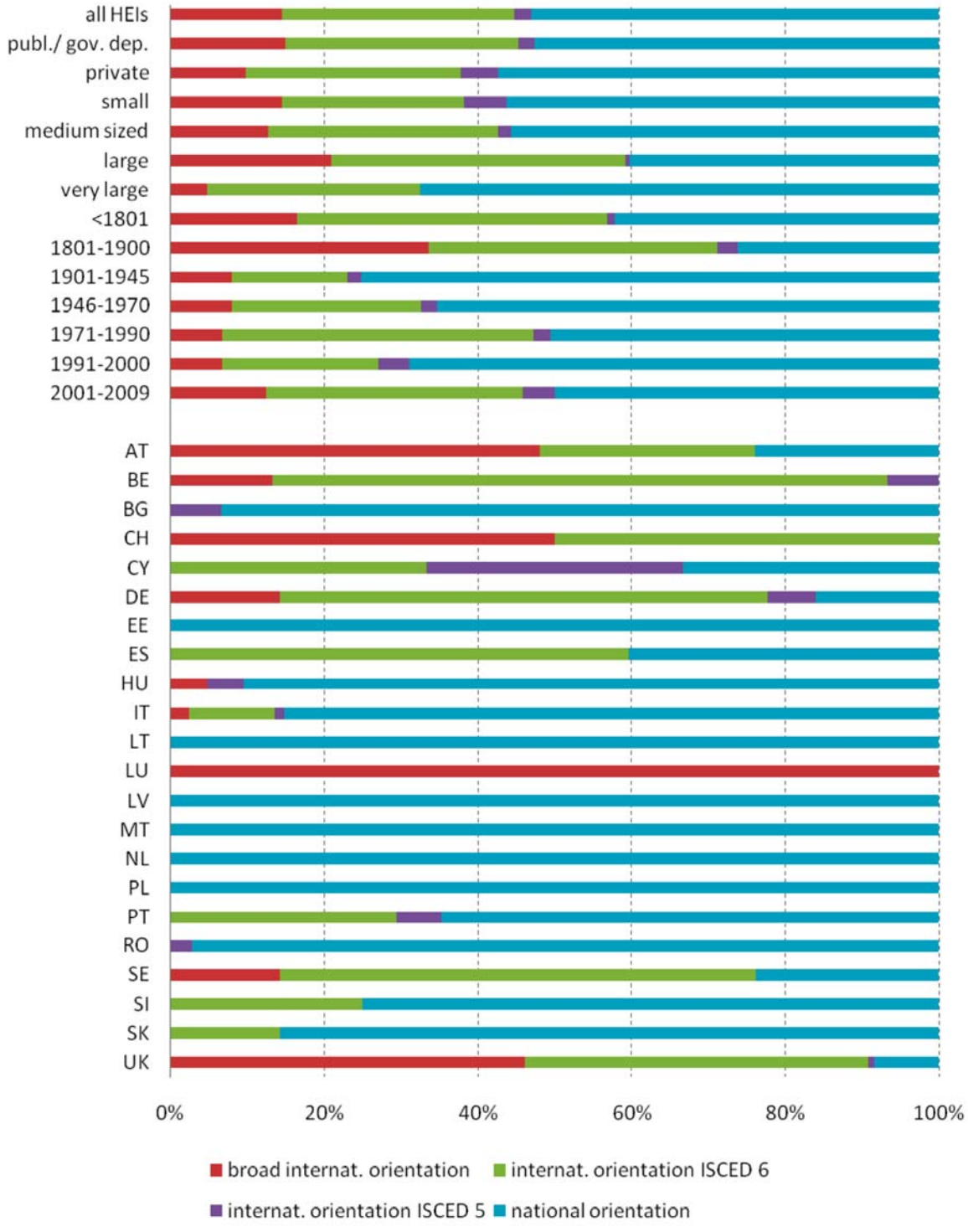


Figure 36. International orientation of research active European higher education institutions by category and country

Source: EUMIDA dataset 2010 excluding Denmark, France (no data available), Czech Republic, Finland, Greece, Ireland, Netherlands, Norway (limited data availability/ comparability)

7.5 Subject mix

The availability of data on students by field of education permits a large scale examination of the problem of subject mix. It is well known that differences in the profile of educational fields are associated to large differences in cost structures, outputs, and various measures of productivity. The investigation of these differences will require a dedicated research effort.
As a preliminary step, we suggest a characterization in four categories:

- Generalist
- Technical universities
- Specialist in Social sciences
- Specialist in Humanities and Arts.

This characterization follows an extensive cluster analysis (not shown here) we performed with various techniques and is quite robust. The variables used for clustering are always the shares of ISCED 5 students in each field of education.
Generalist universities cover several disciplines, none of which has an overwhelming share of students. Technical universities are focused on Engineering, often with a single or a few minor fields, such as Business and economics, or Natural Sciences. Specialists in Social sciences include the Business schools and specialised universities in Law and, to a lesser extent, Political sciences or administration. Finally, Specialist in Humanities and Arts include schools dedicated to music, art, drama, as well as all disciplines in Humanities.
Unfortunately, this analysis cannot be carried out for all countries, since several of them provided incomplete data.
Data show that the dominant model in the European higher education landscape is the generalist one, largely beyond 50% of the total. Among the institutions awarding the doctorate, the share of Generalists exceeds 70%. On the other hand, among the non-doctorate institutions, one can find 20% of Specialists in Humanities and Art, 20% in Social Sciences, and almost 40% Technical schools.
A somewhat related difference can be found between private and public institutions: among the private the Generalists are slightly more than 40% and the Specialists in Social sciences (mainly Business schools and Law schools) exceed 30%.
It can be said that private and non-doctorate institutions exploited the rigidity of the dominant generalist model of the public university in order to identify niches of unfilled opportunities, mainly for teaching purposes.
In fact, generalists are the dominant model for old institutions (more than 80% for those established before 1801) and also for those created in the 19th century. Technical schools have been created with more frequency in the 1971-1990 period, while Specialists in Social sciences flourished after 2000. In the institutions created in the last decade, Generalists are

roughly 40%, and a similar share is represented by Specialists in Social sciences. The transition in Eastern European countries may explain part of this pattern.

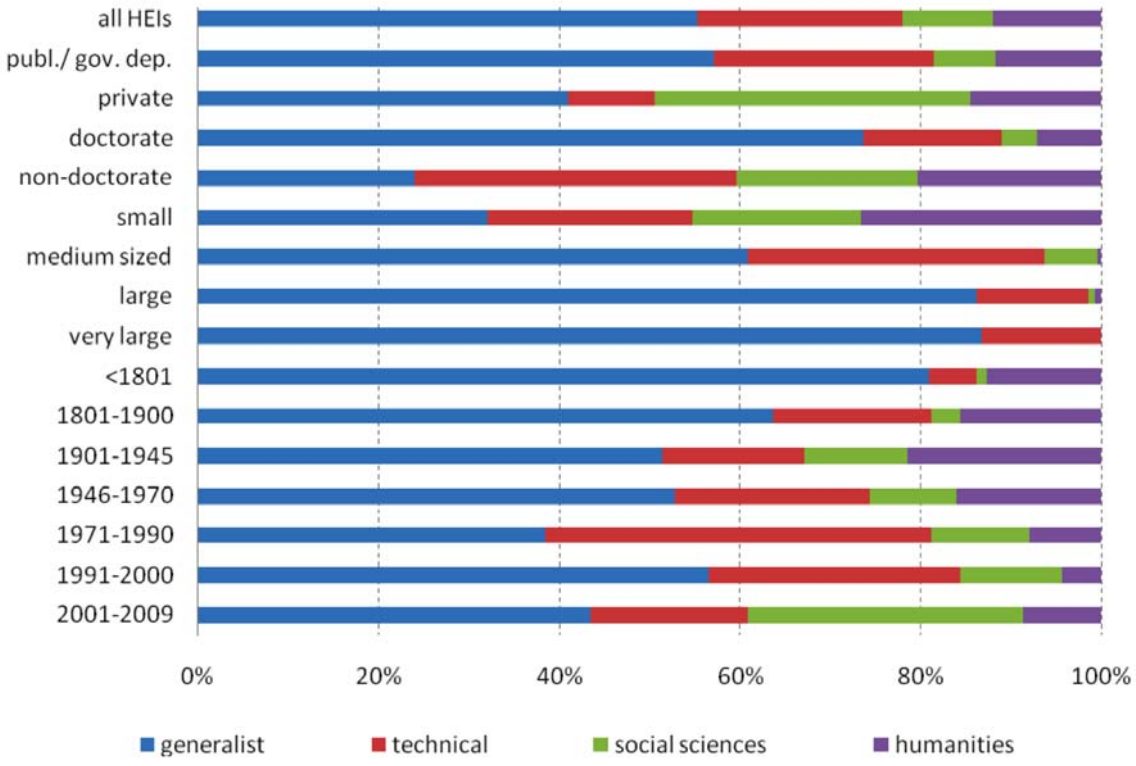


Figure 37. Subject mix of research active European higher education institutions by category

Source: EUMIDA dataset 2010 excluding Denmark, France (no data available), Austria, Cyprus, Czech Republic, Estonia, Finland, Greece, Hungary, Ireland, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Poland, Portugal, Romania (limited data availability/ comparability).

Table 49. Subject mix based on staff by fields of science

All institutions

	%Natural Sciences	%Engineering Technology	%Medical Sciences	%Agricultural Sciences	%Social Sciences	%Humanities
average	13,7%	16,3%	12,2%	2,4%	28,5%	26,9%
median	11,7%	6,9%	0,6%	0,0%	21,2%	15,9%
min	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
max	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%

Generalist

	%Natural Sciences	%Engineering Technology	%Medical Sciences	%Agricultural Sciences	%Social Sciences	%Humanities
average	18,7%	8,8%	21,4%	3,3%	23,7%	24,0%
median	18,4%	6,4%	18,3%	0,0%	21,2%	19,2%
min	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
max	100,0%	75,0%	100,0%	100,0%	74,3%	83,2%

Technical

	%Natural Sciences	%Engineering Technology	%Medical Sciences	%Agricultural Sciences	%Social Sciences	%Humanities
Average	14,2%	49,6%	1,3%	2,3%	25,6%	7,1%
Median	12,9%	44,9%	0,0%	0,0%	23,7%	4,2%
Min	0,0%	20,0%	0,0%	0,0%	0,0%	0,0%
Max	48,2%	100,0%	20,9%	34,9%	73,5%	31,3%

Social sciences

	%Natural Sciences	%Engineering Technology	%Medical Sciences	%Agricultural Sciences	%Social Sciences	%Humanities
average	1,7%	0,4%	0,9%	1,1%	94,0%	1,9%
median	0,0%	0,0%	0,0%	0,0%	100,0%	0,0%
min	0,0%	0,0%	0,0%	0,0%	56,6%	0,0%
max	25,0%	11,1%	14,9%	29,0%	100,0%	20,5%

Humanities and Art

	%Natural Sciences	%Engineering Technology	%Medical Sciences	%Agricultural Sciences	%Social Sciences	%Humanities
average	0,2%	0,8%	0,0%	0,0%	0,6%	98,5%
median	0,0%	0,0%	0,0%	0,0%	0,0%	100,0%
min	0,0%	0,0%	0,0%	0,0%	0,0%	80,5%
max	5,9%	19,5%	1,9%	0,0%	11,3%	100,0%

7.6 PhD intensity

The share of PhD students out of total number of students gives a proxy of the research orientation of universities. We again limit this analysis to the subset of research active institutions, leaving aside a small number of PhD awarding units that however do not meet the criteria for research activity.

Following the convention used in the Carnegie Classification, we distinguish between universities with a share of PhD students below or beyond 5%. This is a conservative definition, however, since data shows that the most active universities have a share that largely exceeds this threshold. Using this definition, it turns out that around 40% of research active universities have high intensity of PhD students, only 5& have zero intensity, and the remaining group lies in the middle.

Further disaggregation of data might show more subtle differences.

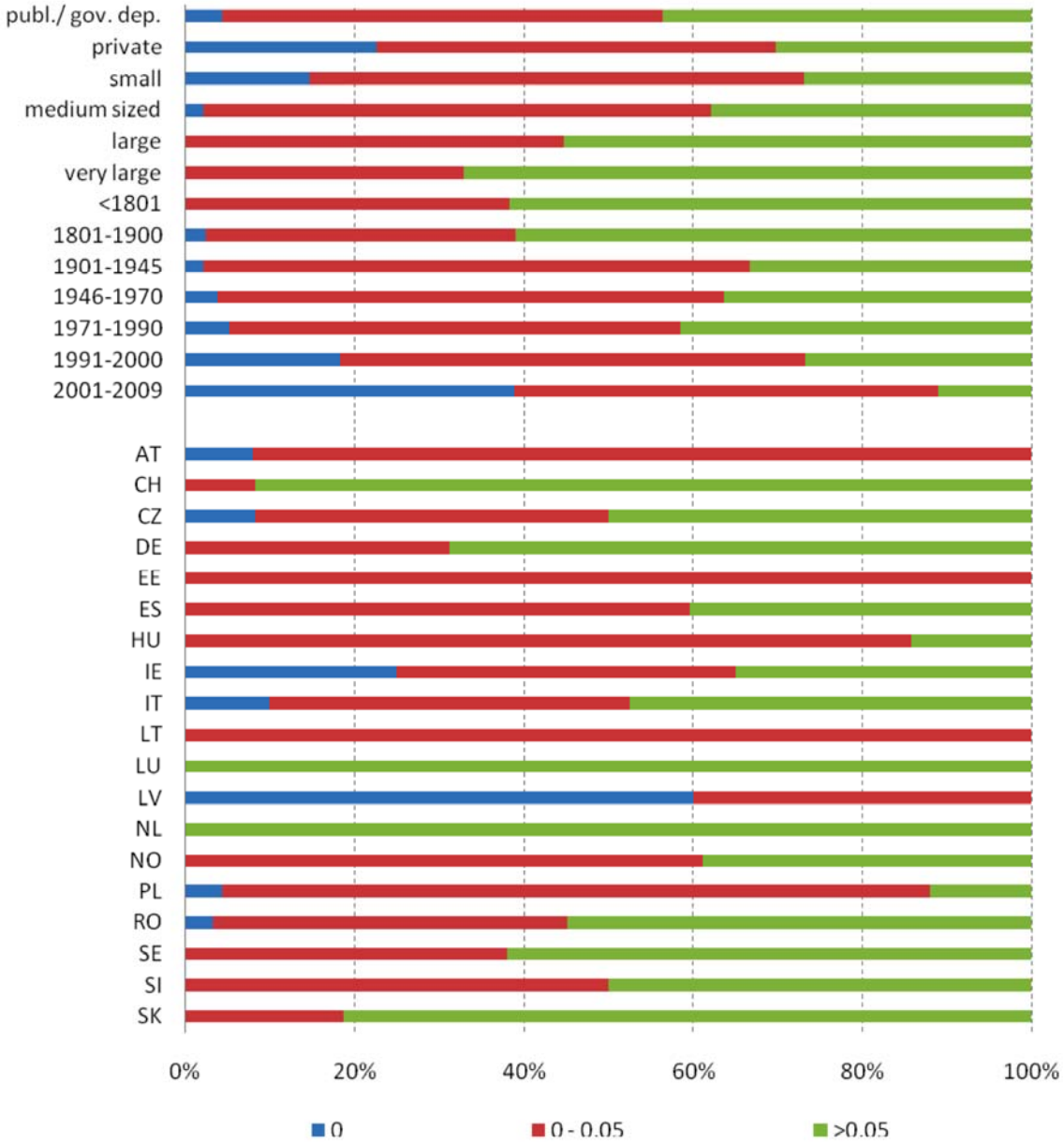


Figure 38. PhD intensity of research active European higher education institutions by category

Source: EUMIDA dataset 2010 excluding Denmark, France (no data available), Belgium, Bulgaria, Cyprus, Finland, Greece, Malta, Portugal, United Kingdom (limited data availability/ comparability).

8. Conclusions and recommendations

8.0 Introduction

The main goal of the EUMIDA project was to test the feasibility of a regular data collection of microdata on higher education institutions in all EU-27 Member States plus Norway and Switzerland. This Report comments on data availability, confidentiality, and on the resources needed for a full-scale exercise. Its main achievement is to have demonstrated that in all countries there actually exists a core set of data that shares the following features:

- Follows the definitions laid down in the UNESCO-OECD-EUROSTAT (UOE) Manual
- Is routinely collected by the National Statistical Authorities (NSAs)
- Does not raise significant confidentiality issues
- Can be disaggregated at the level of individual units in a smooth way.

Another main achievement is to have highlighted the key conceptual, methodological, and practical issues associated to a full-scale data collection on outputs, as was experimentally done in Data Collection 2.

In this scenario, it was not expected that such a project would deliver policy implications *per se*. However, the overall process of collaboration with the NSAs, the setting up of a network of national experts in support to the study team, and a review of the main issues encountered and dealt with during the project, have led the consortium to propose a number of recommendations. These have both a practical, short-term impact, but also some far reaching implications.

8.1 Recommendations on publication of data

The EUMIDA project has collected evidence in Data Collection 1 showing that the core set is complete (by country and by variable) and does not raise significant confidentiality issues. As a result, it would be feasible for the European Commission to make Data Collection 1 public, following a procedure of formal authorization from the NSAs.

A link to the website of the Commission, or a separate website might be created for this purpose.

According to our recommended procedure, the Commission might collect from the NSAs, for each DC1 variable, the authorization to act on either of the following options:

a Publish the data with the name of the institution displayed

b Publish the data without the name of the institution, but only with an identification number associated to it (i.e. anonymously)

c Don't publish the data at all.

The resulting dataset would be subject to a principle of variable geometry, which is a reasonable goal in a field that has historically been characterized by absolute lack of data at

micro level. We anticipate that no serious confidentiality issues would be raised for technical reasons.

With respect to Data Collection 2 the situation looks more problematic, given the limited number of countries and/or variables in the overall coverage spectrum.

Due to several methodological problems still pending, which have been illustrated at great length in this Report, we suggest adopting a more gradual approach here. This would require initiating a wide debate with academia, administrators, and policy makers, as well as the scientific communities interested in higher education, science and technology, and public policy. It is important that indicators would not only be validated from a statistical point of view, but also appreciated through the practical implications, both on analysis and policy. This activity should mostly be focused on DC1 and will also have to address the issue of the perimeter of research active institutions, being one of the key indicators, now taking the form of a dummy, associated to all institutions in the collection. In parallel, further methodological and feasibility work should be done on the variables included in DC2, in the perspective of a future utilization.

8.2 Recommendations on regular data collection

As stipulated earlier, the EUMIDA study has shown that a regular data collection is feasible, does not encounter methodological obstacles that could not be addressed in a reasonable way, and should not prove too costly terms of the resources required.

In light of the above, we propose that EUROSTAT builds on the work begun by EUMIDA and the FESUR Task Force and engages in a process with all the NSAs to define the procedures and methodological requirements for a regular data collection. The EUMIDA Handbook, which is annexed to this Report, could prove useful as a starting point for discussing definitions and data collection procedures.

8.3 Recommendations on statistical capacity building

In order to engage in a regular data collection, it would also be required that NSAs strengthen their capacities at national level in the field of statistics on higher education institutions.

This would imply a joint management of different statistical Manuals, namely the UOE on higher education and the Frascati and Oslo Manuals on research and development. Thus, some level of integration would also be needed between units that are responsible for the production of data having different methodological traditions.

8.4 Recommendations on data on funding and expenditure

Data on financial flows (funding and expenditure) is not systematically available at individual level and suffers from lack of standardization, despite the existence of common definitions

with respect to R&D. Thus, it is recommended that further work is carried out at both NSAs and EUROSTAT level, in order to improve the quality and coverage of this data.
In particular, data on R&D expenditure of higher education institutions is critical information required to analyze their research activities. This evidence is routinely collected in the R&D statistics on the basis of a well-defined methodology that is presented in detail in the Frascati manual. However, experience from the EUMIDA data collection shows that only in about half of the Member States could this data effectively be used at the institutional level, since it is based on regular surveys of staff; in the other countries, national conversion rates are used, some of them based on old surveys, and thus their quality and level of detail are not sufficiently good. This experience matches information already available at EUROSTAT and OECD, which questions whether R&D data from higher education institutions are always of sufficient quality. OECD already signalled their interest in promoting improvements in this area, which are however also dependent on the propensity of Member States to invest additional resources in R&D data collection.
We thus propose a) that EUROSTAT cooperates actively with OECD on improving the quality of data collection in this area, and b) that the Commission actively encourages Member States initiatives to improve their data collection on R&D in higher education institutions.

8.5 Recommendations on further feasibility studies

8.5.1 Publications

The EUMIDA Consortium decided to commit to a feasibility-in-the-feasibility study, using its own budget resources to fund a pilot exercise on bibliometric indicators for universities. The study has been based on the SCOPUS database and has examined a non-random sample of 50 units, most of which are not included in the rankings of top European universities (CWTS, or Shanghai).
The pilot has shown that:

- There exists a production of publications also in higher education institutions that do not reach the visibility level needed to enter into rankings;
- The overall bibliometric profile may change significantly according to the database used. In particular, SCOPUS has a better coverage of some fields of science, which may have an impact due to the heterogeneity in subject mix across institutions;
- The task of matching affiliations in bibliometric databases to names of institutions is easy to tackle in the cases where there is only one university per city, while it becomes more complex (up to one month of effort) for large cities and metropolitan areas.

Overall, the effort needed to build up a full-scale repository of all names of affiliations in the SCOPUS bibliometric database with a matching to the official names of institutions in the

EUMIDA Census is in the order of a budget of Euro 200,000. We consider this a very small effort for a result that might change radically the way in which European HEIs are perceived and evaluated.
We thus suggest that the Commission might consider a new feasibility study with the goal of creating such a repository.

8.5.2 Patents

The issue of measurement of academic patents is well known and the EUMIDA Consortium shares the view that data should not refer to patents for which the assignee is a university.
The methodology developed by Lissoni and colleagues proved to be successful. A limitation to be considered is the cost of this procedure. Another valuable input is the analysis of affiliations of patents carried out by experts at Katholieke University of Leuven (KUL– B), who developed data production methods for harmonized patent statistics aimed in particular at applicants' sectorial allocation.
The Fraunhofer Institute at Kalrsruhe worked with an alternative methodology, based on the automatic matching between lists of authors in bibliometric databases and lists of inventors in patent databases. This approach has produced very significant results with considerably less effort.
We suggest that the Commission might consider a feasibility study on the potential adoption of this methodology to build up a repository of matching of affiliations.

8.5.3 Webometrics

The EUMIDA consortium has worked with CSIS on the use of webometrics information. The potential of such information to examine the visibility of universities on the web has been evaluated in extremely positive terms. However, the relative role of visibility in research or society, and of the web pages containing teaching materials, has not yet been disentangled.
We recommend further research on this issue, after which the inclusion of webometric information in a regular data collection might become a credible option to consider.

8.6 Recommendations on diffusion

We anticipate that the publication of data will raise a number of debates and policy oriented discussions. In fact, we consider the use of indicators not for a top down, authoritative elimination of debates, but rather as tools for discussion and opinion formulation at theoretical and practical level.
Accordingly we would recommend building a comprehensive overview of indicators, to open a forum of stable interaction with stakeholders and policy makers on their definition, assessment and operationalisation. We would suggest that the overall scientific community

of higher education scholars, economists and sociologists of education and research, as well as social scientists, should be involved in such an initiative. The network of national experts convened by EUMIDA might also be a resource for enhancing and qualifying the debate.

ANNEXES

1 Bibliography

2+5 EUMIDA Handbook on Definitions, Acronyms, Data Sources, Surveys and Methodologies

3+4 Data Collection 1 & 2 - Main tables and Data values for the figures

6 Notes on Statistical Validity

7 Principal Participants to the Study

8 List of the Higher Education Institutions included in the EUMIDA database