U-Multirank

Design and Testing the Feasibility of a Multidimensional Global University Ranking

Final Report

Frans van Vught & Frank Ziegele (eds.)

Consortium for Higher Education and Research Performance Assessment

CHERPA-Network

June 2011

CONTRACT - 2009 -1225 /001 -001

The CHERPA Network

CENTRE FOR HIGHER
EDUCATION DEVELOPMENT

Center for
Higher Education
Policy Studies

In cooperation with

U-Multirank Project team

Project leaders
Frans van Vught (CHEPS) *
Frank Ziegele (CHE) *
Jon File (CHEPS) *

Project co-ordinators
Maarja Beerkens-Soo (CHEPS)
Elisabeth Epping (CHEPS) *

Research co-ordinators
Gero Federkeil (CHE) *
Frans Kaiser (CHEPS) *

Research team
Sonja Berghoff (CHE)
Uwe Brandenbrug (CHE)
Julie Callaert (INCENTIM) *
Koenraad Debackere (INCENTIM)
Ghislaine Filliatreau (OST) *
Wolfgang Glänzel (INCENTIM)
Ben Jongbloed (CHEPS) *
Bart van Looy (INCENTIM)
Suzy Ramanana-Rahary (OST)
Isabel Roessler (CHE) *
Françoise Rojouan (OST)
Robert Tijssen (CWTS) *
Phillip Vidal (OST)
Martijn Visser (CWTS)
Don Westerheijden (CHEPS) *
Erik van Wijk (CWTS)
Michel Zitt (OST)

International expert panel
Nian Cai Liu (Shanghai Jiao Tong University)
Simon Marginson (Melbourne University)
Jamil Salmi (World Bank)
Alex Usher (IREG)
Marijk van der Wende (OECD/AHELO)
Cun-Mei Zhao (Carnegie Foundation)

* Contributing authors: U-Multirank final report

Table of Contents

Tables

Figures

Preface

On 2 June 2009 the European Commission announced the launching of a feasibility study to develop a multi-dimensional global university ranking.

Its aims were to "look into the feasibility of making a multi-dimensional ranking of universities in Europe, and possibly the rest of the world too". The Commission believes that accessible, transparent and comparable information would make it easier for students and teaching staff, but also parents and other stakeholders, to make informed choices between different higher education institutions and their programmes. It would also help institutions to better position themselves and improve their quality and performance.

The Commission pointed out that existing rankings tend to focus on research in "hard sciences" and ignore the performance of universities in areas like humanities and social sciences, teaching quality and community outreach. While drawing on the experience of existing university rankings and of EU-funded projects on transparency in higher education, the new ranking system should be:

- **multi-dimensional:** covering the various missions of institutions, such as education, research, innovation, internationalisation, and community outreach;
- **transparent:** it should provide users with a clear understanding of all the factors used to measure performance and offer them the possibility to consult the ranking according to their needs;
- **global:** covering institutions inside and outside Europe (in particular those in the US, Asia and Australia).

The project would consist of two consecutive parts:

- In a first phase running until the end of 2009 the consortium would **design** a multi-dimensional ranking system for higher education institutions in consultation with stakeholders.
- In a second phase ending in June 2011 the consortium would **test the feasibility** of the multi-dimensional ranking system on a sample of no less than 150 higher education and research institutions. The sample would focus on the disciplines of engineering and business studies and should have a sufficient geographical coverage (inside and outside of the EU) and a sufficient coverage of institutions with different missions.

In undertaking the project the consortium was greatly assisted by four groups that it worked closely with:

- An **Advisory Board** constituted by the European Commission as the project initiator which included not only representatives of the Directorate General: Education and Culture but other experts drawn from student organisations, employer organisations, the OECD, the Bologna Follow-up Group and a number of Associations of Universities. The Advisory Board met seven times over the course of the project.
- An **international expert panel** composed of six international experts in the field of mapping, ranking and transparency instruments in higher education and research. The international panel was consulted at key decision making moments in the project.
- Crucially, given the user-driven nature of the new transparency instrument designed within the project, interested and committed **stakeholder representatives** met with the project team over the life of the project. The stakeholder consultations provided vital input on the relevance of potential performance dimensions and indicators, on methods of presenting the rankings in an informative and user-friendly format, and on different models for implementing the new transparency instrument. Stakeholder workshops were held four times during the project with an average attendance of 35 representatives drawn from a wide range of organisations including student bodies, employer organisations, rectors' conferences, national university associations and national representatives.
- The consortium members benefitted from a strong **network of national higher education experts** in over 50 countries who were invaluable in suggesting a diverse group of institutions from their countries to be invited to participate in the pilot study.

This is the **Final Report** of the multi-dimensional global university ranking project. Readers interested in a fuller treatment of many of the topics covered in this report are referred to the project web-site (www.u-multirank.eu) where the project's three **Interim Reports** can be found.

The web-site also includes a 30 page **Overview** of the major outcomes of the project.

Executive Summary

The need for a new transparency tool in higher education and research

The project encompassed the **design** and **testing** of a **new transparency tool** for higher education and research. More specifically, the focus was on a transparency tool that will enhance our understanding of the **multiple performances** of different higher education and research institutions across the diverse range of activities they are involved in: higher education and research institutions are multi-purpose organisations and different institutions focus on different blends of purposes and associated activities.

Transparency is of major importance for higher education and research worldwide which is increasingly expected to make a crucial contribution to the innovation and growth strategies of nations around the globe. Obtaining valid information on higher education within and across national borders is critical in this regard, yet higher education and research systems are becoming more complex and – at first sight – less intelligible for many stakeholders. The more complex higher education systems become, the more sophisticated our transparency tools need to be. Sophisticated tools can be designed in such a way that they are user-friendly and can cater to the different needs of a wide variety of stakeholders.

An enhanced understanding of the diversity in the profiles and performances of higher education and research institutions at a national, European and global level requires a new ranking tool. Existing international transparency instruments do not reflect this diversity adequately and tend to focus on a single dimension of university performance – research. The new tool will promote the development of diverse institutional profiles. It will also address most of the major shortcomings of existing ranking instruments, such as language and field biases, the exaggeration of small differences in performance and the arbitrary effects of indicator weightings on ranking outcomes.

We have called this new tool *U-Multirank* as this stresses three fundamental points of departure: it is **multi-dimensional**, recognising that higher education institutions serve multiple purposes and perform a range of different activities; it is a ranking of university **performances** (although not in the sense of an aggregated league table like other global rankings); and it is **user-driven** (as a stakeholder with particular interests, **you** are enabled to rank institutions with comparable profiles according to the criteria important to you).

The design and key characteristics of U-Multirank

On the basis of a carefully selected set of design principles we have developed a new international ranking instrument that is **user-driven, multi-dimensional** and **methodologically robust**. This new **on-line** instrument enables its users first to identify institutions that are **sufficiently comparable** to be ranked and, second, to design a personalised ranking by selecting the indicators of particular relevance to them. U-Multirank enables such comparisons to be made both at the level of institutions as a whole and in the broad disciplinary fields in which they are active. The integration of the already designed and tested **U-Map** classification tool into U-Multirank enables the creation of the user-selected groups of sufficiently comparable institutions. This two-step approach is completely new in international and national rankings.

On the basis of an extensive **stakeholder consultation process** (focusing on relevance) and a thorough **methodological analysis** (focusing on validity, reliability and feasibility)**, U-Multirank** includes a range of indicators that will enable users to compare the performance of institutions across five dimensions of higher education and research activities:

- Teaching and learning
- Research
- Knowledge transfer
- International orientation
- Regional engagement

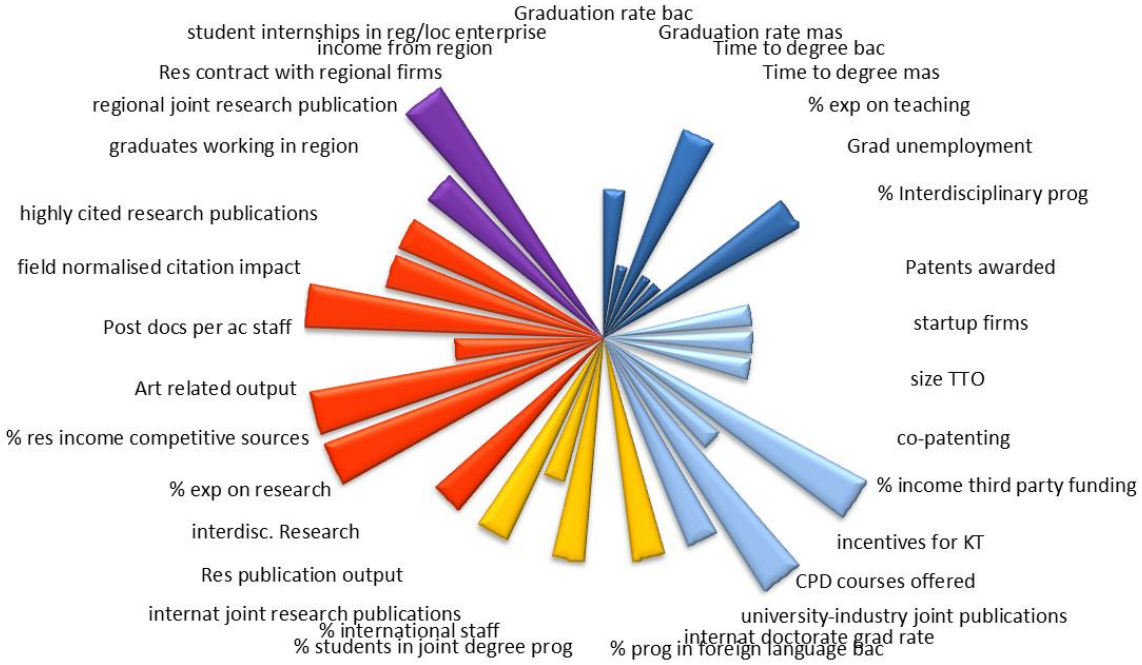
On the basis of data gathered on these indicators across the five performance dimensions, U-Multirank could provide its users with the on-line functionality to create two general types of rankings:

- **Focused institutional rankings**: rankings on the indicators of the five performance dimensions at the level of institutions as a whole
- **Field-based rankings**: rankings on the indicators of the five performance dimensions in a specific field in which institutions are active

U-Multirank would also include the facility for users to create institutional and field **performance profiles** by including (not aggregating) the indicators within the five dimensions (or a selection of them) into a multi-dimensional performance chart. At the institutional level these take the form of 'sunburst charts' while at the field level these are structured as 'field-tables'.

In the sunburst charts, the performance on all indicators at the institutional level is represented by the size of the rays of the 'sun': a larger ray means a higher performance on that indicator. The colour of a ray reflects the dimension to which it belongs. The sunburst chart gives an impression 'at a glance' of the performance of an institution, without unwarranted aggregation of information into composite indicators.

Figure 1: Sunburst representation of an institutional performance profile



In the field based table below relative performance is indicated by a coloured circle. A green circle indicates that the score of the institution on that indicator is in the top group, a red circle indicates that the performance is in the bottom group, and a yellow circle means that performance is somewhere in the middle. The user may sort the institutions on all of the indicators presented. In addition the users are given the opportunity to choose the indicators on which they want to rank the institutions selected. This personalised interactive ranking table reflects the user driven nature of U-Multirank.

Table 1: Performance at the field level

	Teaching & Learning			Research			Knowledge transfer			international orientation			Regional engagement		
	student staff ratio	graduation rate	qualification of academic staff	research publication output	external research income	citation index	% income third party funding	CPD courses offered	startup firms	international academic staff	% international students	joint international publ.	graduates working in the region	student internships in	regional co-publication
Institution 4	🟢	🟡	🟡	🟢	🔴	🟡	🔴	🟡	🟡	🟢	🟡	🔴	🟢	🔴	🟡
Institution 8	🟢	🟡	🟡	🟡	-	🟡	🟡	🔴	🟡	🔴	🟡	-	🟡	🟢	🟢
Institution 3	🟡	🟢	🟢	🟡	🟢	🟡	🟡	🟢	🟡	🟡	🟢	🟡	🟡	-	🟡
Institution 5	🟡	🟢	🟢	🔴	🟢	🟢	🟡	🟢	🔴	🟡	🔴	🟢	🟡	🟢	-
Institution 1	🟡	-	-	🟡	🟢	🟢	🟢	🟡	🟢	🟡	🟡	🟢	🟢	🟡	🟢
Institution 9	🟡	🔴	🔴	🟡	🔴	🔴	🔴	🟡	🔴	🟢	🟡	🔴	🔴	🟡	🔴
Institution 7	🟡	🟡	🔴	🟡	-	🔴	🟢	-	🟢	-	🟡	🟡	🟡	🔴	🟡
Institution 2	🔴	🟡	🟡	🟢	🟢	🟢	-	🔴	🔴	🟡	🔴	🟡	🔴	🔴	🔴
Institution 6	🔴	🔴	🟡	🔴	-	🔴	🟢	🟡	🟢	🔴	🟢	🟢	🟢	🟡	🟢

In order to be able to apply the principle of **comparability** we have integrated the existing transparency tool – the U-Map classification – into U-Multirank. U-Map has been designed, tested and is now being implemented through a series of projects also supported by the European Commission. It is a user driven higher education mapping tool that allows users to select comparable institutions on the basis of 'activity profiles' generated by the U-Map tool. These activity profiles reflect the diverse activities of different higher education and research organisations using a set of dimensions similar to those developed in U-Multirank. The underlying indicators differ as U-Map is concerned with understanding the mix of activities an institution is engaged in (*what it does*), while U-Multirank is concerned with an institution's performance in these activities (*how well it does what it does*). Integrating U-Map into U-Multirank enables the creation of user-selected groups of sufficiently comparable institutions that can then be compared in focused institutional or field based rankings.

The findings of the U-Multirank pilot study

U-Multirank was tested in a pilot study involving 159 higher education institutions drawn from 57 countries: 94 from within the EU; 15 from other European countries; and 50 from outside Europe.

The pilot test demonstrated that multi-dimensional institutional and field level ranking is certainly possible in terms of the development of feasible and relevant indicators. It also showed the value of multi-dimensionality with many institutions and faculties performing very differently across the five dimensions and their underlying indicators. The multi-dimensional approach makes these diverse performances transparent.

While indicators on teaching and learning, research, and internationalisation proved largely unproblematic, in some dimensions (particularly knowledge transfer and regional engagement) and with some concepts (such as graduate employability and non-traditional research output) feasible indicators are more difficult to develop.

In terms of the potential **level of institutional interest** in participating in the new transparency tool, the results of the pilot study are positive. In broad terms, half of the institutions invited to participate in the pilot study agreed to do so. Given that a significant number of these institutions (32%) were from outside Europe and it is clear that U-Multirank is a Europe-based project, this represents a strong expression of interest. Institutions with a wide range of activity profiles demonstrated their interest in participating.

The pilot study suggests that a multi-dimensional ranking would be feasible in Europe. However, difficulties with the availability and comparability of information mean that it would be unlikely to achieve extensive coverage levels across the globe in the short-term. There are however clear signals that there would be significant continuing interest from outside Europe from institutions wishing to benchmark themselves against European institutions.

In terms of the feasibility of "up-scaling" a pilot project of 150 institutions to one including ten or twenty times that number and extending its field coverage from three to around fifteen major disciplinary fields, the pilot study suggests that while this will bring significant logistical, organisational and financial challenges, there are no inherent features of U-Multirank that rule out the possibility of such future growth.

In summary, the pilot study demonstrates that in terms of the feasibility of the dimensions and indicators, potential institutional interest in participating, and

operational feasibility we have developed a U-Multirank 'Version 1.0' that is ready to be implemented in European higher education and research as well as for institutions and countries outside Europe that are interested in participating.

The further development and implementation of U-Multirank

The outcomes of the pilot study suggest some clear next steps in the further development of U-Multirank and its implementation in Europe and beyond. These include:

- **The refinement of U-Multirank instruments:** Some modifications need to be made to a number of indicators and to the data gathering instruments based on the experience of the pilot study. Crucially, the on-line ranking tool and user-friendly modes of presenting ranking outcomes need to be technically realised.
- **Roll-out of U-Multirank across European countries:** Given the need for more transparent information in the emerging European higher education area all European higher education and research institutions should be invited to participate in U-Multirank in the next phase.
- Many European stakeholders are interested in assessing and comparing European higher education and research institutions and programmes globally. **Targeted recruitment of relevant peer institutions from outside Europe** should be continued in the next phase of the development of U-Multirank.
- **Developing linkages with national and international data-bases.**
- **The design of specific authoritative rankings:** Although U-Multirank has been designed to be user driven, this does not preclude the use of the tool and underlying database to produce authoritative "expert" institutional and field based rankings for particular groups of comparable institutions on dimensions particularly relevant to their activity profiles.

In terms of the organisational arrangements for these activities we favour a further two year project phase for U-Multirank. In the longer term on the basis of a detailed analysis of different organisational models for an institutionalised U-Multirank our strong preference is for an **independent non-profit organisation** operating with multiple sources of funding**.** This organisation would be independent both from higher education institutions (and their associations) and from higher education governance and funding bodies. Its non-commercial character will add legitimacy as will external supervision via a Board of Trustees.

1 Reviewing current rankings

1.1 Introduction

This chapter summarises the findings of our extensive analysis of currently existing transparency tools. Readers interested in a more comprehensive treatment of this topic are referred to the project's first interim report (January 2010)[1]. First, we present our argument for user-driven rankings being an epistemic necessity. Secondly, we present the results of the extensive review of the different transparency tools - quality assurance, classifications, and rankings - from the point of view of the information they could deliver to assist different stakeholders in their different decisions regarding higher education and research institutions. Thirdly, we consider the impact of current rankings - both negative and (potentially) positive. Finally, we identify some indications for better practice, both theoretically inspired and based on existing good practices.

1.2 User-driven rankings as an epistemic necessity

Each observation of reality is theory-driven: every observation of a slice of reality is influenced by the conceptual framework that we use to address it. In the scientific debate, this statement is accepted at least since Popper's work (Popper, 1980): he has shown abundantly that theories are 'searchlights' that cannot encompass all of reality, but necessarily highlight only certain aspects of it. He also showed that scientific knowledge is 'common sense writ large' (Popper, 1980, p. 22), meaning that the demarcation between common sense and scientific knowledge is that the latter has to be justified rationally: scientific theories are logically coherent sets of statements, which moreover are testable to show if they are consistent with the facts.

Failing conceptual frameworks or scientific theories, many areas of life (such as for instance sports) have been organised with (democratic) forums that have been accepted as authorities to set rules. The conceptual frameworks behind sports league tables are usually well-accepted: rules of the game define who the winner is and how to make a league table out of that. Yet those rules have been designed by humans and may be subject to change: in the 1980s-1990s football associations went from 2 points for winning a match to 3 points, changing the tactics in the game (more attacks

[1] See www.u-multirank.eu

late in a drawn match), changing the league table outcome to some extent, and sparking off debates among commentators of the sport for and against the new rule.[1]

In university rankings, the rules of the ranking game are equally defined by humans, because there is no scientific theory of what is 'the best university'. But different to sports, there are no officially recognised bodies that are accepted as authorities that may define the rules of the game. There is no understanding, in other words, that e.g. the Shanghai ranking is simply a game that is as different from the Times Higher game as rugby is from football. And that the organisation making up the one set of rules and indicators has no more authority than the other to define a particular set of rules and indicators. The issue with the some of the current university rankings is that they tend to be presented as if their collection of indicators did reflect *the* quality of the institution; they have the pretension, in that sense, of being guided by a (non-existent) theory of the quality of higher education.

We do not accept that position. Our alternative to assuming an unwarranted position of authority is to reflect critically on the different roles that higher education and research institutions have for different groups of stakeholders, to define explicitly our conceptual framework regarding the different functions of higher education institutions, and in turn to derive sets of indicators from this framework. And then to present the information encapsulated in those indicators in such a transparent way that the actual users of rankings can make their own decisions about what counts for them as being best for their purpose(s), resulting in their own specific and time-dependent rankings. In this sense, we want to democratise rankings in higher education and research. Based on the epistemological position that any choice of sets of indicators is driven by their makers' conceptual frameworks, we suggest a user-driven approach to rankings. Users and stakeholders themselves should be enabled to decide which indicators they want to select to create the rankings that are relevant to their purposes. We want to give them the tools and the information to make their own decisions.

1.3 Transparency, quality and accountability in higher education

It is widely recognized that although the current transparency tools—especially university league tables—are controversial, they seem to be here to stay, and that especially global university league tables have a great impact on decision-makers at all levels in all countries, including in universities (Hazelkorn, 2011). They reflect a growing international competition among universities for talent and resources; at the same time they reinforce competition by their very results. On the positive side they

[1] http://en.wikipedia.org/wiki/Three_points_for_a_win

urge decision-makers to think bigger and set the bar higher, especially in the research universities that are the main subjects of the current global league tables. Yet major concerns remain as to league tables' methodological underpinnings and to their policy impact on stratification rather than on diversification of mission.

Let us first define the main concepts that we will be using in this report. Under *vertical stratification* we understand distinguishing higher education and research institutions as 'better' or 'worse' in prestige or performance; *horizontal diversification* is the term for differences in institutional missions and profiles. Regarding the different instruments, *transparency tool* is the most encompassing term in our use of the word, including all the others; it denotes all manners of providing insight into the diversity of higher education. Transparency tools are instruments that aim to provide information to stakeholders about the efforts and performance of higher education and research institutions. A *classification* is a systematic, nominal distribution among a number of classes or characteristics without any (intended) order of preference. Classifications give descriptive categorizations of characteristics intending to focus on the efforts and activities of higher education and research institutions, according to the criterion of similarity. They are eminently suited to address horizontal diversity. *Rankings* are hierarchical categorizations intended to render the outputs of the higher education and research institutions according to the criterion of best performance. Most existing rankings in higher education take the form of a league table. A *league table* is a single-dimensional, ordinal list going from 'best' to 'worst', assigning to the entities unique, discrete positions seemingly equidistant from each other (from 1 to, e.g., 500). Transparency tools are related to quality assurance processes. *Quality assurance,* evaluation or accreditation, also produces information to stakeholders (review reports, accreditation status) and in that sense helps to achieve transparency. As the information function of quality assurance is not very elaborate (usually only informing if basic quality, e.g. the accreditation threshold, has been reached) and as quality assurance is too ubiquitous to allow for an overview on a global scale in this report, in the following we will focus on classifications and rankings. Let us underline here, though, that rankings and classifications on the one hand and quality assurance on the other play complementary roles.

In the course of our project, we undertook an extensive review of the different transparency tools, - quality assurance, classifications and rankings - from the point of view of which information they could deliver to assist users in their different decisions regarding higher education and research institutions. The results of this extensive review are presented in the project's interim report (CHERPA-Network, 2010).

Table 1-1: Classifications and rankings considered in U-Multirank

Type	Name
Classifications	• Carnegie classification (USA) • U-Map (Europe)
Global League Tables and Rankings	• Shanghai Jiao Tong University's (SJTU) Academic Ranking of World Universities (ARWU) • Times Higher Education (Supplement) (THE) • QS (Quacquarelli Symonds Ltd) Top Universities • Leiden Ranking
National League Tables and Rankings	• US News & World Report (USN&WR; USA) • National Research Council (USA) PhD programs • Times Good Education Guide (UK) • Guardian ranking (UK) • Forbes (USA) • CHE Das Ranking / University Ranking (CHE; Germany) • Studychoice123 (SK123; the Netherlands)
Specialized League Tables and Rankings	• Financial Times ranking of business schools and programmes (FT; global) • BusinessWeek (business schools, USA + global) • The Economist (business schools; global)

The major dimensions along which we analysed the classifications, rankings and league tables included:

- Level: e.g. institutional vs. field-based
- Scope: e.g. national vs. international
- Focus: e.g. education vs. research
- Primary target group: e.g. students vs. institutional leaders vs. policy-makers
- Methodology and producers: which methodological principles are applied and what sources of data are used and by whom?

We concluded from our review that different rankings and classifications use different methodologies, implying but often not explicating different conceptions of quality of higher education and research. Most are presented as league tables; especially the most influential ones, the global university rankings are all league tables. The relationship of indicators collected and their weights in calculating the league table rank of an institution are not based on explicit let alone scientifically justifiable conceptual frameworks. Moreover, indicators often are distant proxies to quality. It seems that availability of quantitative data has precedence over their

validity and reliability. In recent years, probably due to the influence of widely-published guidelines such as the Berlin Principles of ranking (International Ranking Expert Group, 2006) and of recent initiatives such as the U-Map classification (van Vught et al., 2010) and even already anticipating the current U-Multirank project, the situation has begun to change: ranking producers are becoming more explicit and reflective about their methodologies and underlying conceptual frameworks. Increasingly also, web tools of rankings begin to include some degree of interactivity and choice for end users.

Notwithstanding differences in methodologies and their recent improvements, by and large the well-known criticisms of rankings remain valid (Dill & Soo, 2005; Usher & Savino, 2006; Van Dyke, 2005) and are borne out in more recent criticisms (Hazelkorn, 2011; Rauhvargers, 2011), which can be summarised as a set of methodological problems of rankings:

- The problem of unspecified target groups: different users have different information needs while most rankings give only a single ranking
- The problem of ignoring diversity within higher education and research institutions: many rankings are at the institutional level, ignoring that education and research performances may differ much across programmes and departments
- The problem of narrow range of dimensions: most rankings focus on indicators of research, ignoring education and other functions of higher education and research institutions (practice-oriented research, innovation, 'third mission')
- The problem of composite overall indicators: most rankings add or average the indicators into a single number, ignoring that they are about different dimensions and sometimes use different scales
- The problem of league tables: most rankings are presented as league tables, assigning each institution at least those in the top-50, unique places, suggesting that all differences in indicators are valid and of equal weight (equidistant positions).
- The problem of field and regional biases in publication and citation data: many rankings use bibliometric data, ignoring that the available international publication and citation databases mainly cover peer reviewed journal articles, while that type of scientific communication is prevalent only in a narrow set of disciplines (most natural sciences, some fields in medicine) but not in many others (engineering, other fields in medicine and natural sciences, humanities and social sciences)

- The problem of unspecified and volatile methodologies: in many cases, users cannot obtain the information necessary to understand how rankings have been made; moreover, especially commercial publishers of rankings have been accused of changing their ranking methodologies to ensure changes in the top-10 to boost sales rather than to focus on stability and comparability of rankings from year to year.

At the same time, our review uncovered some good practices in the world of rankings, some of which have a beneficial influence on others active in this realm, while practically all informed the design of U-Multirank. We already mentioned some of them. The full list includes:

- The Berlin Principles on Ranking of Higher Education Institutions (International Ranking Expert Group, 2006), which define sixteen standards and guidelines to make rankings transparent, user-oriented (clear about their target group), and focusing on performance
- Rankings for students such as those of CHE and Studychoice123, which have a clear focus based on a single target group, and which are presented in a very interactive, user-oriented manner enabling custom-made rankings rather than dictating a single one
- Focused institutional rankings, in particular the Leiden ranking of university research, also with a clear focus, not pretending to assess all-round quality, and with a transparent methodology
- Qualifications frameworks and Tuning Educational Structures, showing that at least qualitatively it is possible to define performances regarding student learning thus strengthening the potential information base for other dimensions than fundamental research
- Comparative assessment of higher education student's learning outcomes (AHELO): this feasibility project of the OECD to develop a methodology extends the focus on student learning introduced by Tuning and by national qualifications frameworks into an international comparative assessment of undergraduate students, much like PISA does for secondary school pupils.
- Recent reports on rankings such as the report of the Assessment of University-Based Research Expert Group (AUBR Expert Group, 2009) which defined a number of principles for sustainable collection of research data, such as purposeful definition of the units or clusters of research, attention to the use of non-obtrusive measurement e.g. through digital repositories of publications, leading to a matrix of data that could be used in different constellations to respond to different scenarios (information needs).

Our review also included an extensive survey of the indicators used in current classifications and rankings, to ensure that in the development of the set of indicators for U-Multirank we would not overlook any dimensions, data sources or lessons learned about data and data collection. The results of this part of the exercise will be reflected in the next chapters.

We realise explicitly that there is no neutral measurement of social issues; each measurement—the operationalization of constructs, the definition of indicators, and the selection of data sources—depends on the interest of research and the purpose of the measurement. International rankings in particular should be aware of possible biases and be precise about their objectives and how those are linked to the data they gather and display.

The global rankings that we studied limit their interest to several hundred pre-selected universities, estimated to be no more than 1% of the total number of higher education institutions worldwide. The criteria used to establish a threshold generally concern the research output of the institution; the amount of research output, in other words the institution's visibility in research terms, is generally seen as a prerequisite for being ranked on a global scale. Although it could be argued that world-class universities may act as role models (Salmi, 2009), the evidence that strong institutions inspire better performance across whole higher education systems is so far mainly found in the area of research rather than that of teaching (Sadlak & Liu, 2007) if there are positive system-wide spill-overs at all (Cremonini, Benneworth & Westerheijden, 2011).

From our overview of the indicators used in the main global university rankings (summarised in Table 1-2) we concluded that they focus indeed heavily on research aspects of the higher education institutions (research output, impact as measured through citations, and reputation in the eyes of academic peers) and that efforts to include the education dimension remain weak and use distant 'proxy' indicators. Similarly, the EUA in a recent overview also judged that these global rankings provide an 'oversimplified picture' of institutional mission, quality and performance, as they focus mainly on indicators related to the research function of universities (Rauhvargers, 2011).

Table 1-2: Indicators and weights in global university rankings

	HEEACT 2010	ARWU 2010	THE 2010	QS 2011	Leiden Rankings 2010
Research output	Articles past 11 years (10%) and last year (10%)	Articles published in Nature and Science (20%) [Not calculated for institutions specialized in humanities and social sciences]	Research income (5.25%) Ratio public research income / total research income (0.75%) Papers per staff member (4.5%)		Number of publications (P)
Research impact	Citations last 11 years (10%) and last 2 years (10%) Average annual number of citations last 11 years (10%) Hirsch-index last 2 years (20%) Highly-cited papers (15%) Articles last year in high-impact journals (15%)	Articles in Science Citation Index-expanded and Social Science Citation Index (20%)	Citations (normalised average citation per paper) (32.5%)	Citations per faculty member (20%)	Two versions of size-independent, field-normalized average impact ('crown indicator' CPP/FCSm, and alternative calculation MNCS2) Size-dependent 'brute force' impact indicator (multiplication of P with the university's field-normalized average impact): P * CPP/FCSm Citations-per-publication indicator (CPP)
Quality of education		Alumni of an institution winning Nobel Prizes and Fields Medals (10%)	PhDs awarded per staff (6%) Undergraduates admitted per staff (4.5%) Income per staff (2.25%) Ratio PhD awards / bachelor awards (2.25%)	Faculty student ratio (20%)	

	HEEACT 2010	ARWU 2010	THE 2010	QS 2011	Leiden Rankings 2010
Quality of staff		Staff winning Nobel Prizes and Fields Medals (20%) Highly cited researchers in 21 broad subject categories (20%)			
Reputation			Peer review survey (19.5+15=34.5%) International staff score (5%) International students score (5%)	Academic reputation survey (40%) Employer reputation survey (10%)	
General		Sum of all indicators, divided by staff number (10%)	Ratio international mix, staff and students (5%) Industry income per staff (2.5%)	International faculty (5%) International students (5%)	
Website	http://ranking.heeact.edu.tw/en-us/2010/Page/Indicators	http://www.arwu.org/ARWUMethodology2010.jsp	http://www.timeshighereducation.co.uk/world-university-rankings/2010-2011/analysis-methodology.html	http://www.topuniversities.com/university-rankings/world-university-rankings	http://www.socialsciences.leiden.edu/cwts/products-services/leiden-ranking-2010-cwts.html
Notes					There are several rankings, each focusing on one indicator.

A major reason why the current global rankings focus on research data is that this is the only type of data readily available internationally. Potentially, the three main ways of collecting information for use in rankings seem to be:

- Use of statistics from existing databases. National databases on higher education and research institutions cover different information based on national, different definitions of items and are therefore not easily used in cross-national comparisons. International databases such as those of UNESCO, OECD and the EU show those comparability problems but moreover they are focused on the national level and are therefore not useful for institutional or field comparisons.[3] International databases with information at the institutional level or lower aggregation levels are currently available for specific subfields: research output and impact, and knowledge transfer and innovation. Regarding research output and impact, there are worldwide databases on journal publications and citations (the well-known Thomson Reuters and Scopus databases). These databases, after thorough checking and adaptation, are used in the research-based global rankings. Their strengths and weaknesses were mentioned above. Patent databases have not been used until now for global rankings.
- Self-reported data collected by higher education and research institutions participating in a ranking. This source is used regularly though not in all global rankings, due to the lack of externally available and verified statistics (Thibaud, 2009). Self-reported data ought to be externally validated or verified; several methods to that end are available. The drawback is high expense for the ranking organisation and for the participating higher education and research institutions.
- Surveys among stakeholders such as staff members, students, alumni or employers. Surveys are strong methods to elicit opinions such as reputation or satisfaction, but are less suited for gathering factual data. Student satisfaction and to a lesser extent satisfaction of other stakeholders is used in national rankings, but not in existing global university rankings. Reputation surveys are used globally, but have been proven to be very weak cross-nationally (Federkeil, 2009) even if the sample design and response rates were acceptable, which is not often the case in the current global university rankings. Manipulation of opinion-type data has surfaced in surveys for ranking and is hard to uncover or validate externally.

A project closely linked with ours, U-Map, has tested 'pre-filling' higher education institutions' questionnaires, i.e. data available in national public sources are entered into

[3] The beginnings of European data collection as in the EUMIDA project may help to overcome this problem for the European region in years to come.

the questionnaires sent to higher education institutions for data gathering. This should reduce the effort required from higher education institutions and give them the opportunity to verify the 'pre-filled' data as well. The U-Map test with 'pre-filling' from national data sources in Norway appeared to be successful and resulted in a substantial decrease of the burden of gathering data at the level of higher education institutions.

1.4 Impacts of current rankings

According to many commentators, impacts of rankings on the sector are rather negative: they encourage wasteful use of resources, promote a narrow concept of quality, and inspire institutions to engage in 'gaming the rankings'. As will be shown near the end of this section, a well-designed ranking can have a positive effect on the sector, encouraging higher education and research institutions to improve their performance. Impacts may affect amongst other things:

- Student demand. There is evidence that student demand and enrolment in study programmes rises after positive statements in national, student-oriented rankings. Both in the US and Europe rankings are not equally used by all types of students (Hazelkorn, 2011): less by domestic undergraduate entrants, more at the graduate and postgraduate levels. Especially at the undergraduate level, rankings appear to be used particularly by students of high achievement and by those coming from highly educated families (Cremonini, Westerheijden & Enders, 2008; Heine & Willich, 2006; McDonough, Antonio & Perez, 1998).
- Institutional management. Rankings strongly impact on the management in higher education institutions. The majority of higher education leaders report that they use potential improvement in rank to justify claims on resources (Espeland & Saunder, 2007; Hazelkorn, 2011). In institutional actions to improve ranking positions, they tend to focus on targeting the indicators in league tables that are most easily influenced, e.g. the institution's branding, institutional data and choice of publication language (English) and channels (journals counted in the international bibliometric databases). Moreover, there are various examples of cases in which leaders' salary or their positions were linked to their institution's position in rankings (Jaschik, 2007).
- Public policy, in particular public funding. In nations across the globe, global rankings have prompted the desire for 'world-class universities' both as symbols of national achievement and prestige and supposedly as engines of the knowledge economy (Marginson, 2006). It can be questioned if redirecting funds to a small set of higher education and research institutions to make them 'world class' benefits the whole higher education system; research on this question is lacking until now.

- The higher education 'reputation race'. The reputation race (van Vught, 2008) implies the existence of an ever-increasing search by higher education and research institutions and their funders for higher positions in the league tables. In Hazelkorn's survey of higher education institutions, 3% were ranked first in their country, but 19% wanted to get to that position (Hazelkorn, 2011). The reputation race has costly implications. The problem of the reputation race is that the investments do not always lead to better education and research, and that the resources spent might be more efficiently used elsewhere. Besides, the link between quality in research and quality in teaching is not particularly strong (see Dill & Soo, 2005).
- Quality of higher education and research institutions. Rankings' incomplete conceptual and indicator frameworks tend to get rooted as definitions of quality (Tijssen, 2003). This standardization process is likely to reduce the horizontal diversity in higher education systems.
- 'Matthew effect'. As a result of the vertical differentiation, rankings are likely to contribute to wealth inequality and expanding performance gaps among institutions (van Vught, 2008). This is sometimes called a 'Matthew effect' (Matthew 13:12), i.e. a situation where already strong institutions are able to attract more resources from students (e.g. increase tuition fees), government agencies (e.g. research funding), and third parties, and thereby to strengthen their market position even further.
- 'Gaming the results'. Institutional leaders are under great pressure to improve their institution's position in the league tables. In order to do so, these institutions sometimes may engage in activities that improve their positions in rankings but which may have negligent or even harmful effects on their performance in core activities.

Most of the effects discussed above are rather negative to students, institutions and the higher education sector. The problem is not so much the existence of rankings as such, but rather that many existing rankings are flawed and create dysfunctional incentives. If a ranking would be able to create useful incentives, it could be a powerful tool for improving the performance in the sector. Well-designed rankings may be used as a starting point for internal analysis of strengths and weaknesses. Similarly, rankings may provide useful stimuli to students to search for the best-fitting study programmes and to policy-makers to consider where in the higher education system investment should be directed for the system to fulfil its social functions optimally. The point of the preceding observations was not that all kinds of stakeholders react to rankings, but that the current rankings and league tables seem to invite overreactions on too few dimensions and indicators.

1.5 Indications for better practice

Our critical review also resulted in some indications for a better practice, both theoretically inspired and looking at existing good practices. They are as follows:

- As suggested in the Berlin Principles, rankings should explicitly define and address target groups, as indicators and the way to present results have to be focused.
- Rankings and quality assurance mechanisms are complementary instruments. Rankings represent an external, quantitative view on institutions from a transparency perspective; traditional instruments of internal and external quality assurance are aiming at institutional accountability and quality enhancement. Rankings are not similar to quality assurance instruments but they may help to ask the right questions for processes of internal quality enhancement.
- For some target groups, in particular students and researchers, information has to be field-based; for others, e.g. university leaders and national policy-makers, information about the higher education institution as a whole has priority (related to the strategic orientation of institutions); a multi-level set of indicators must reflect these different needs.
- In rankings comparisons should be made between higher education and research institutions of similar characteristics, leading to the need for a pre-selection of a set of more or less homogeneous institutions. Rankings that include very different profiles of higher education and research institutions are non-informative and misleading.
- Rankings have to be multidimensional. The various functions of higher education and research institutions for a heterogeneity of stakeholders and target groups can only be adequately addressed in a multidimensional approach.
- There are neither theoretical nor empirical reasons for assigning fixed weights to individual indicators to calculate a composite overall score; within a given set of indicators the decision about the relative importance of indicators should be left to the users.
- International rankings have to be aware of potential biases of indicators; aspects of international comparability therefore have to be an important issue in any ranking. .
- Rankings should not use league tables from 1 to *n* but should differentiate between clear and robust differences in levels of performance. The decision about an adequate number of 'performance categories' has to be taken with regard to the number of institutions included in a ranking and the distribution of data.
- Rankings have to use multiple databases to bring in different perspectives on institutional performance. As much as possible available data sources should be used, but currently their availability is limited. To create multidimensional

rankings, gathering additional data from the institutions is necessary. Therefore, the quality of the data collection process is crucial.

- In addition rankings should be self-reflexive with regard to potential unintended consequences and undesirable/perverse effects.
- Involvement of stakeholders in the process of designing a ranking tool and selecting indicators is crucial to keep feedback loops short, so as to avoid misunderstandings and so as to enable a high quality of the designed instruments.
- A major issue is the measures to ensure quality of the ranking process and instruments. This includes statistical procedures as well as the inclusion of the expertise of stakeholders, rankings and indicator experts, field experts (for the field-based rankings) and regional/national experts. A crucial aspect is transparency about the methodology. The basic methodology, the ranking procedures, the data used (including information about survey samples) and the definitions of indicators have to be public for all users. Transparency includes information about the limitations of the rankings.

These general conclusions have been an important source of inspiration for how we designed U-Multirank, a new, global, multidimensional ranking instrument. Based on these conclusions, in the next chapter we will formulate the design principles that have guided the development of this new tool.

2 Designing U-Multirank

2.1 Introduction

Based on the findings of our analyses of the currently existing transparency tools (see chapter 1) this chapter addresses the basic design aspects of a new, multidimensional global ranking tool that we have called 'U-Multirank'. First, we present the general design principles that to a large extent have guided the design process. Secondly, we describe the conceptual framework from which we deduce the five dimensions of the new ranking tool. Finally, we outline a number of methodological choices that have a major impact on the operational design of U-Multirank.

2.2 Design Principles

U-Multirank aims to address the challenges identified as arising from the various currently existing ranking tools. Using modern theories and methodologies of design processes as our base (Bucciarelli, 1994; Oudshoorn & Pinch, 2003) and trying to be as explicit as possible about our approach, we formulated a number of design principles that guided the development of the new ranking tool. The following list contains the basic principles applied when designing and constructing U-Multirank.

- Our fundamental epistemological argument is that as all observations of reality are theory-driven (formed by conceptual systems) an 'objective ranking' cannot be developed (see chapter 1). Every ranking will reflect the normative design and selection criteria of its constructors.
- Given this epistemological argument, our position is that rankings should be based on the interests and priorities of their users: rankings should be **user-driven**. This principle "democratizes" the world of rankings by empowering potential users (or categories of users) to be the dominant actors in the design and application of rankings rather than rankings being restricted to the normative positions of a small group of constructors. Different users and stakeholders should be able to construct different sorts of rankings. (This is one of the *Berlin Principles*).
- Our second principle is **multidimensionality**. Higher education and research institutions are predominantly multi-purpose, multiple-mission organizations undertaking different mixes of activities (teaching and learning, research, knowledge transfer, regional engagement, and internationalization are five major categories that we have identified; see the following section). Rankings should reflect this multiplicity of functions and not focus on one function (research) to the virtual exclusion of all else. An obvious corollary to this principle is that

institutional performance on these different dimensions should never be aggregated into a composite overall ranking.

- The next design principle is **comparability**. In rankings, institutions and programs should only be compared when their purposes and activity profiles are sufficiently similar. Comparing institutions and programs that have very different purposes is worthless. It makes no sense to compare the research performance of a major metropolitan research university with that of a remotely located University of Applied Science; or the internationalization achievements of a national humanities college whose major purpose is to develop and preserve its unique national language with an internationally orientated European university with branch campuses in Asia.
- The fourth principle is that higher education rankings should reflect the **multilevel nature of higher education**. With very few exceptions, higher education institutions are combinations of faculties, departments and programs of varying strength. Producing only aggregated institutional rankings disguises this reality and does not produce the information most valued by major groups of stakeholders: students, potential students, their families, academic staff and professional organizations. These stakeholders are mainly interested in information about a particular field. This does not mean that institutional-level rankings are not valuable to other stakeholders and for particular purposes. The new instrument should allow for the comparisons of comparable institutions at the level of the organization as a whole and also at the level of the disciplinary fields in which they are active.
- Finally we include the principle of **methodological soundness**. The new instrument should refrain from methodological mistakes such as the use of composite indicators, the production of league tables and the denial of contextuality. In addition it should minimise the incentives for strategic behaviour on the part of institutions to 'game the results'.

These principles underpin the design of U-Multirank, resulting in a user-driven, multidimensional and methodologically robust ranking instrument. In addition, U-Multirank aims to enable its users to identify institutions and programs that are sufficiently comparable to be ranked, and to undertake both institutional and field level analyses.

A fundamental question regarding the design of any transparency tool has to do with the choice of the 'dimensions': on which subject(s) will the provision of information focus? What will be the topics of the new ranking tool?

We take the position that any process of collecting information is driven by a – more or less explicit – conceptual framework. Transparency tools should clearly show what these

conceptual frameworks are and how they have played a role in the selection of the broader categories of information on which these tools are focused.

For the design of U-Multirank we specify our own conceptual framework in the following section.

2.3 Conceptual framework

A meaningful ranking requires a conceptual framework in order to decide on its content categories. We call these categories the 'dimensions' of the new ranking tool. We found a number of points of departure for a general framework for studying higher education and research institutions in the higher education literature. Four different conceptual perspectives have been combined in this approach.

First, a common point of departure is that processing knowledge is the general characteristic of higher education and research institutions (Clark 1983; Becher and Kogan 1992). 'Processing' can be the discovery of new knowledge as in research, or its transfer to stakeholders outside the higher education and research institutions (knowledge transfer) or to various groups of 'learners' (education). Of course, a focus on the overall objectives of higher education and research institutions in the three well-known primary processes or functions of 'teaching and learning, research, and knowledge transfer' is a simplification of the complex world of higher education and research institutions. These institutions are, in varying combinations of focus, committed to the efforts to discover, conserve, refine, transmit and apply knowledge (Clark 1983). But the simplification helps to encompass the wide range of activities in which higher education and research institutions are involved. The three functions are a useful way to describe conceptually the general purposes of these institutions and therefore are the underlying three dimensions of our new ranking tool.

The second conceptual assumption is that the performance of higher education and research institutions may be directed at different 'audiences'. In the current higher education and research policy area, two main general audiences have been prioritised, the first through the international orientation of higher education and research institutions. This emphasises how these institutions are seen as society's portals to the globalised world (both 'incoming' influences and 'outgoing' contributions to the international discourse). At the same time, the institutions' engagement with the region can be distinguished. Here the emphasis is on the involvement with and impact on the region in which a higher education institution operates. In reality these 'audiences' are of course often combined in the various activities of higher education and research institutions.

It is understood that the functions higher education and research institutions fulfil for international and regional audiences are manifestations of their primary processes, i.e. the three functions of education, research and knowledge transfer mentioned before. What we mean by this is that there may be educational elements, research elements and knowledge transfer elements to the international orientation. Similarly, regional engagement may be evident in an institution's education, research and knowledge transfer activities. International and regional orientation are two further dimensions of the multidimensional ranking.

The term 'processing' used above points to the third main conceptual assumption, namely the major stages in any process of creation or production: input, throughput (or the process in a narrow sense) and its results, which can be subdivided into immediate outputs and further-reaching impacts. A major issue in higher education and research institutions, as in many social systems, has been that the transformation from inputs to performances is not self-evident. One of the reasons why there is so much criticism of league tables is exactly the point that from similar sets of inputs, different higher education and research institutions may reach quite different types and levels of performance.

We make a general distinction between the 'enabling' stages of the overall creation stages on the one hand and the 'performance' stages on the other. The enabling stages consist of the inputs and processes of creation/production processes while the performance stages include their outputs and impacts. We have used the distinction of the various stages of a creation/production process to further elaborate the conceptual framework for the selection of indicators in the new ranking instrument.

A fourth assumption refers to the different stakeholders or users of rankings. Ranking information is produced to inform users about the value of higher education and research, which is necessary as it is not obvious that they are easily able to take effective decisions without such information. (Higher) education is not an ordinary 'good' for which the users themselves may assess the value *a priori* (using, e.g., price information). Higher education is to be seen as an experience good (Nelson 1970): the users may assess the quality of the good only while or after 'experiencing' it (i.e. the higher education program), but such 'experience' is *ex post* knowledge. It is not possible for users to know beforehand whether the educational program meets their standards or criteria. *Ex ante* they only can refer to the perceptions of previous users. Some even say that higher education is a credence good (Dulleck and Kerschbamer 2006): the value of the good cannot be assessed while experiencing it, but only (long) after. If users are interested in the value added of a degree program on the labor market, information on how well a class is taught is not relevant. They need information on how the competences acquired during higher education will improve their position on the career or social ladder. So

stakeholders and users have to rely on information that is provided by a variety of transparency tools and quality assessment outcomes. However, different users require different types of information.

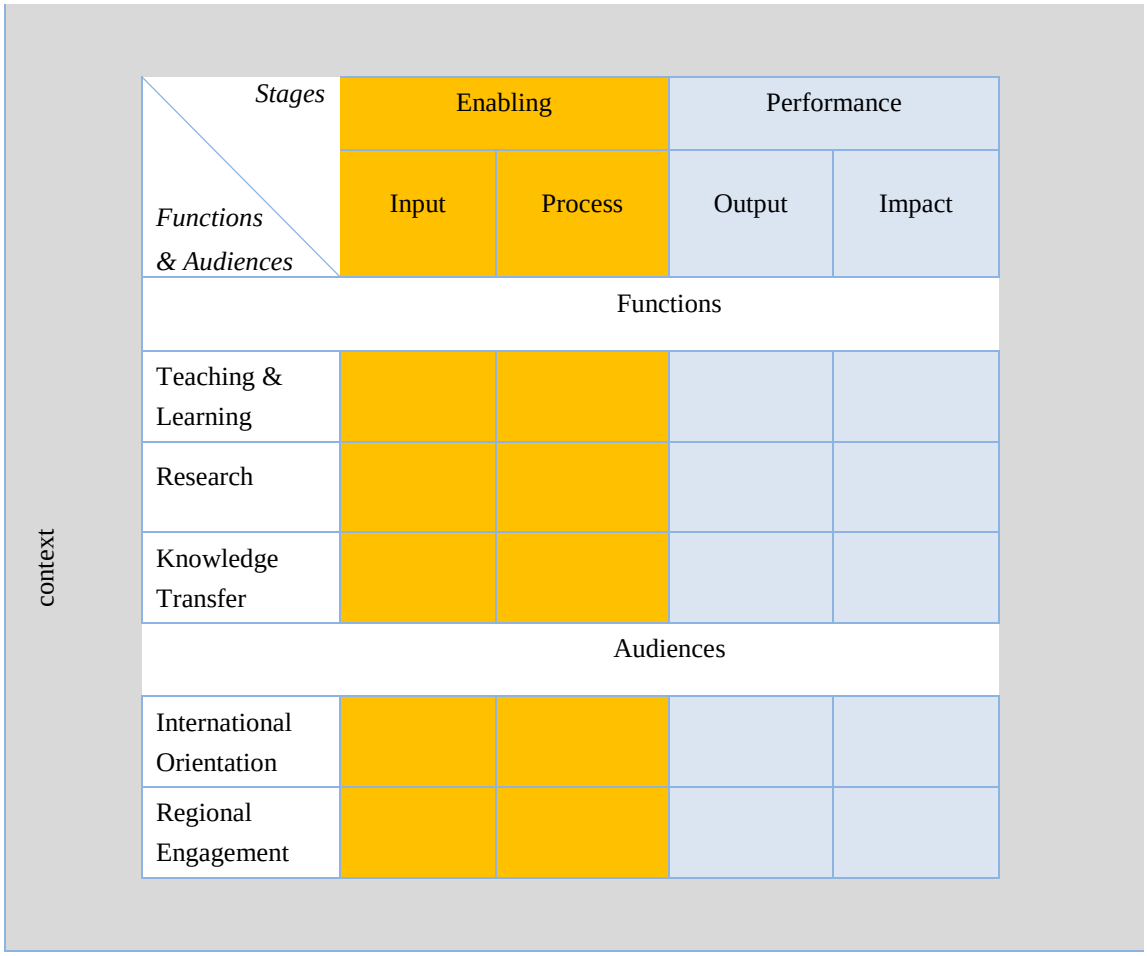
Some users are interested in the overall performance of higher education and research institutions (e.g. policy-makers) and for them the internal processes contributing to performance are of less interest. The institution may well remain a 'black box' for these users. Other stakeholders (students and institutional leaders are prime examples) are interested precisely in what happens inside the box. For instance, students may want to know the quality of teaching in the field in which they are interested. They may want to know how the program is delivered, as they may consider this as an important aspect of their learning experience and their time in higher education (consumption motives). Students might also be interested in the long-term impact of taking the program as they may see higher education as an investment and are therefore interested in its future returns.

Users engage with higher education for a variety of reasons and therefore will be interested in different dimensions and performance indicators of higher education institutions and the programs they offer. Rankings must be designed in a balanced way and include relevant information on the various stages of knowledge processing which are relevant to the different stakeholders and their motives for using rankings.

The conceptual grid shown below must be applied twice: once to the institution as a whole and once at the field level, and it has to accommodate interest in both performance and (to a lesser extent) process. For different dimensions (research, teaching & learning, knowledge transfer) and different stakeholders/users the relevance of information about different aspects of performance may vary.

The result of this elementary conceptual framework is a matrix showing the types of indicator that could be used in rankings and applied at both institutional and field levels. Filtering higher education and research institutions into homogeneous groups requires contextual information rather than only the input and process information that is directly connected with enabling the knowledge processes. Contextual information for higher education and research institutions relates to their positioning in society and specific institutional appearances. It describes the conditions in which the primary processes of education, research and knowledge transfer operate. A substantial part of the relevant context is captured by applying another multidimensional transparency tool (U-Map) in pre-selecting higher education and research institutions (see below). Additional context information may be needed to allow for the valid interpretation of specific indicators by different stakeholders.

Table-2-1: Conceptual grid U-Multirank



Using this conceptual framework we have selected the following five dimensions as the major content categories of U-Multirank:

- Teaching & Learning
- Research
- Knowledge Transfer
- International Orientation
- Regional Engagement

In chapter 3 we will discuss the various indicators to be used in these five dimensions.

An important factor in the argument against rankings and league tables is the fact that often their selection of indicators is guided primarily by the (easy) availability of data rather than by relevance. This often leads to an emphasis on indicators of the enabling stages of the higher education production process, rather than on the area of performance, largely because governance of higher education and research institutions has concentrated traditionally on the bureaucratic (in Weber's neutral sense of the word) control of inputs: budgets, personnel, students, facilities, etc. Then too, inputs and

processes can be influenced by managers of higher education and research institutions. They can deploy their facilities for teaching, but in the end it rests with the students to learn and, after graduation, work successfully with the competencies they have acquired. Similarly, higher education and research institution managers may make facilities and resources available for research, but they cannot guarantee that scientific breakthroughs are 'created'. Inputs and processes are the parts of a higher education and research institution's system that are best documented. But assessing the performance of these institutions implies a more comprehensive approach than a narrow focus on inputs and processes and the dissatisfaction among users of most current league tables and rankings is because they often are more interested in institutional performance while the information they get is largely about inputs. In our design of U-Multirank we focused on the selection of output and impact indicators. U-Multirank intends to be a multidimensional *performance* assessment tool and thus needs to imply indicators that relate to the performances of higher education and research institutions.

2.4 Methodological aspects

There are a number of methodological aspects that have a clear impact on the way a new, multidimensional ranking tool like U-Multirank can be developed. In this section we explain the various methodological choices made when designing U-Multirank.

2.4.1 Methodological standards

In addition to the content-related conceptual framework, the new ranking tool and its underlying indicators must be based also on methodological standards of empirical research, validity and reliability in the first instance. In addition, because U-Multirank is an international comparative transparency tool, it must deal with the issue of comparability across cultures and countries and finally, in order to become sufficiently operational, U-Multirank has to address the issue of feasibility.

Validity

(Construct) validity refers to the evidence about whether a particular operationalization of a construct adequately represents what is intended by the theoretical account of the construct being measured. When characterizing, e.g. the internationality of a higher education institution, the percentage of international students is a valid indicator only if scores are not heavily influenced by citizenship laws. Using the nationality of the qualifying diploma on entry has therefore a higher validity than using citizenship of the student.

Reliability

Reliability refers to the consistency of a set of measurements or measuring instrument. A measure is considered reliable if, repeatedly applied in the same population, it would

always arrive at the same result. This is particularly an issue with survey data (e.g. among students, alumni, staff) used in rankings. In surveys and with regard to self-reported institutional data, the operationalizing of indicators and formulation of questions requires close attention – in particular in international rankings, where cross-cultural understanding of the questions will be an issue.

Comparability

A ranking is the comparison of institutions and programs using numerical indicators. Hence the indicators and underlying data/measure must be comparable between institutions; they have to measure the same quality in different institutions. In addition to the general issue of comparability of data across institutions, international rankings have to deal with issues of international comparability. National higher education systems are based on national legislation setting specific legal frameworks, including legal definitions (e.g. what/who is a professor). Additional problems arise from differing national academic cultures. Indicators, data elements and underlying questions have to be defined and formulated in a way that takes such contextual variations into account. For example, if we know that doctoral students are counted as academic staff in some countries and as students in others, we need to ask for the number of doctoral students counted as academic staff in order to harmonise data on academic staff (excluding doctoral students).

Feasibility

The objective of U-Multirank is to design a multidimensional global ranking tool that is feasible in practice. The ultimate test of the feasibility of our ranking tool has to be empirical: can U-Multirank be applied in reality and can it be applied with a favourable relation between benefits and costs in terms of financial and human resources? We report on the empirical assessment of the feasibility of U-Multirank in chapter 6 of this report.

2.4.2 User-driven approach

To guide the readers' understanding of U-Multirank, we now briefly describe the way we have methodologically worked out the principle of being user-driven (see section 2.2). We propose an interactive web-based approach, where users will be able to declare their interests in a three step, user-driven process:

1. select a set of institutions or fields in institutions ('units') that are homogeneous on descriptive issues judged by the users to be relevant given their declared interests;

2. choose whether to focus the ranking on higher education and research institutions as a whole (focused institutional rankings) or on fields within these institutions (field-based rankings);

3. select a set of indicators to rank the chosen units. This will result in users creating their own specific and different rankings, according to their needs and wishes, from the entire database.

The first step can be based on the existing U-Map classification tool (see section 2.4.3). We argue that it does not make sense to compare all institutions irrespective of their missions, profiles and characteristics, so a selection of comparable institutions based on U-Map should be the basis for any ranking.

In the second step, the users make their choices regarding the ranking level, i.e. whether a ranking will be created at the institutional level, creating a focused institutional ranking, or at the field level, creating a field-based ranking.

The final step is the selection of the indicators to be used in the ranking. There are two ways to organise this choice process. In the first option, users have complete freedom to select from the overall set of indicators, choosing any indicator, addressing any cell in the conceptual grid. Through this personalised approach the users may find information on those aspects in which they are particularly interested. Compared to existing league tables we see this as one of the advantages of our approach. However this kind of individualised, one-off ranking (which may be different even if the same user applies different indicators) may not be attractive to all types of users, as there is no clear nonrelative result for a particular institution or program. To create a user-friendly instrument, guidance tools to take users through the dataset must be established.

2.4.3 U-Map and U-Multirank

The principle of comparability (see section 2.2) calls for a method that helps us in finding institutions the purposes and activity patterns of which are sufficiently similar in order to enable useful and effective rankings. Such a method, we suggest, can be found in the connection of U-Multirank with U-Map (see www.u-map.eu).

U-Map, being a classification tool, describes ('maps') higher education institutions on a number of dimensions, each representing an aspect of their activities. This mapping produces *activity profiles* of the institutions, displaying what the institutions do and how that compares to other institutions. U-Map can prepare the ground for U-Multirank in the sense that it helps identify those higher education institutions that are comparable and for which, therefore, performance can be compared by means of the U-Multirank ranking tool. A detailed description of the methodology used in this classification can be found on the U-Map website (http://www.u-map.eu/methodology.doc/) and in the final report of the U-Map project, which is available at http://www.u-map.org/U-MAP_report.pdf.

Where U-Map is describing what the institutions do (and thus offers descriptive profiles), U-Multirank focuses on the *performance* aspects of higher education and research institutions. U-Multirank shows how well the higher education institutions are performing in the context of their institutional profile. Thus, the emphasis is on indicators of performance, whereas in U-Map it lies on the *enablers* of that performance – the inputs and activities. Despite the difference in emphasis, U-Map and U-Multirank share the same conceptual model. The conceptual model provides the rationale for the selection of the indicators in both U-Map and U-Multirank, both of which are complementary instruments for mapping diversity, horizontal diversity in classification and vertical diversity in ranking.

2.4.4 Grouping

U-Multirank does not calculate league tables. As has been argued in chapter 1, league table rankings have severe flaws which make them, methodologically speaking, unreliable as transparency tools. As an alternative U-Multirank uses a grouping method. Instead of calculating "exact" league table positions we will assign institutions to a limited number of groups.

Within groups there will be no further differentiation. Between the groups statistical methods guarantee that there is a clear difference between performance levels of different groups. The number of groups should be related to the number of institutions ranked. On the one hand the number of groups should express clear differences of performance; on the other hand the number should not be so low as to be restrictive, with the end result that many institutions end up clustered in one group. Last but not least, the number of groups and the methods for calculating the groups must be clear and comprehensible to users.

2.4.5 Design context

In this chapter we have described the general aspects of the design process regarding U-Multirank. We have indicated our general design principles; we have described the conceptual framework from which the five dimensions of U-Multirank are deduced, and we have outlined a number of methodological approaches to be applied in U-Multirank. Together these elements form the design context from which we have constructed U-Multirank.

The design choices made here are in accordance with both the Berlin Principles and the recommendations by the Expert Group on the Assessment of University-based Research. The Berlin Principles[4] emphasize (a.o.) the importance of being clear about the purpose of rankings and their target groups, of recognising the diversity of institutional profiles,

[4] http://www.ireg-observatory.org/index.php?option=com_content&task=view&id=41&Itemid=48

providing users the option to create tailor-made approaches, and of the need to focus on performance rather than on input factors. The AUBR Expert Group[5] (a.o.) underlines the importance of stakeholders' needs and involvement, as well as the principles of purposefulness, contextuality, and multidimensionality of rankings.

Based on our design context, in the following chapters we report on the construction of U-Multirank.

[5] Expert Group on Assessment of University-Based Research (2010), Assessing Europe's University-Based Research, European Commission, DG Research, EUR 24187 EN, Brussels

3 Constructing U-Multirank: Selecting indicators

3.1 Introduction

Having set out the design context for U-Multirank in the previous chapter, we now turn to a major part of the process of constructing U-Multirank: the selection and definition of the indicators. These indicators are assumed to enable us to measure the performances of higher education and research institutions both at the institutional and at the field level, in the five dimensions identified in our conceptual framework (see 2.3): teaching & learning, research, knowledge transfer, international orientation, regional engagement. This chapter provides an overview of the sets of indicators selected for the five dimensions, and describes the selection process. The other important components of the construction process for U-Multirank are the databases and the data collection tools that allow us to actually 'fill' the indicators. These will be discussed further in chapter 4 as we explain the design of U-Multirank in more detail. In chapters 5 and 6 we report on the U-Multirank pilot study during which we analysed the data quality and availability of the various indicators in practice.

3.2 Stakeholders' involvement

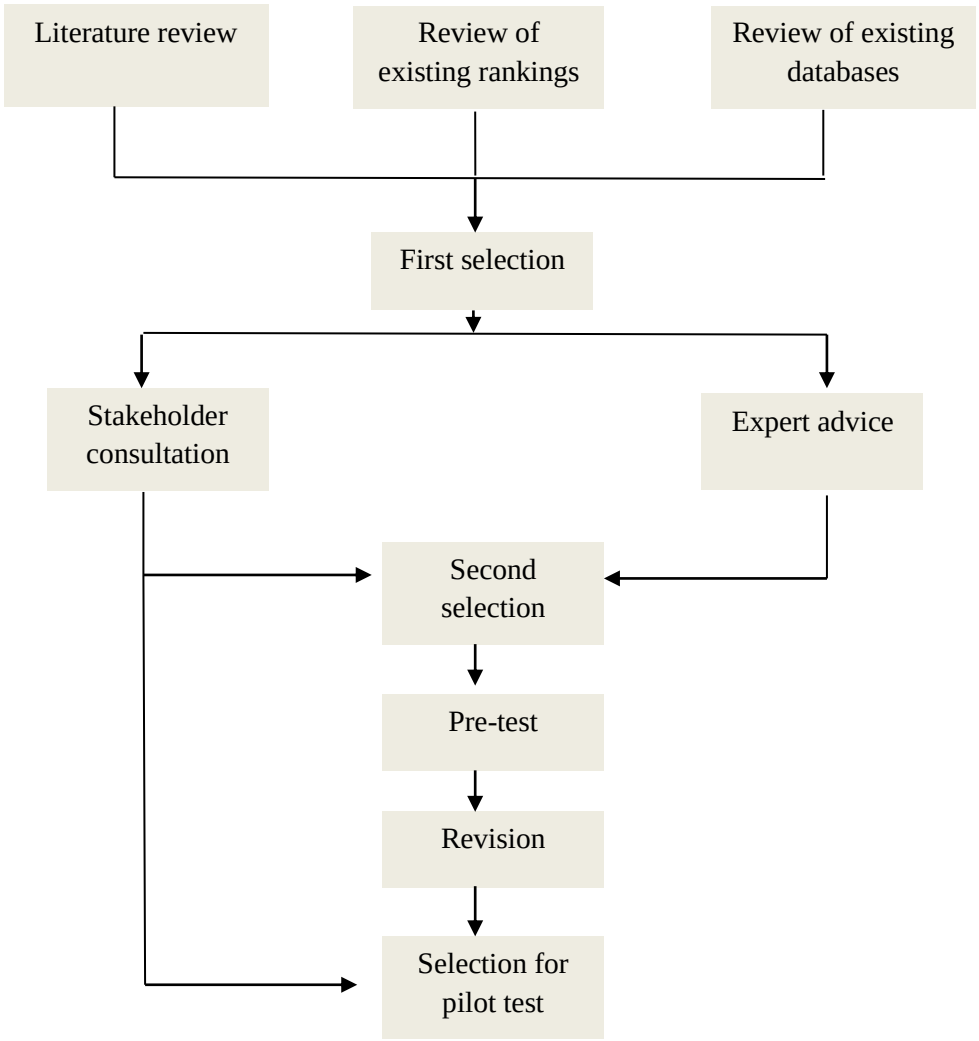
The indicator selection process is illustrated in Figure 3-1. This process is highly stakeholder-driven. Various categories of stakeholders (student organizations, employer organizations, associations and consortia of higher education institutions, government representatives, international organizations) have been involved in an iterative process of consultation to come to a stakeholder-based assessment of the relevance of various indicators. This involvement has been a critical component of our construction process.

The first step in the indicator selection process was a comprehensive inventory of potential indicators from the literature and from existing rankings and databases. This first list was exposed for feedback to stakeholders as well as to groups of specialist experts. Stakeholders were asked to give their views on the relative relevance of various indicators, presented to them as potential items in the five dimensions of U-Multirank (see 3.3). In addition, we invited feedback from international experts in higher education and research and from the Advisory Board of the U-Multirank project.

The information gathered was fed into a second round of consultations with stakeholder organizations. In all some 80 national and international organizations participated in the consultation process. To further support the stakeholder consultation process, an on-line questionnaire was used. Through this process an additional 40 organizations offered their views. To facilitate the consultation process we showed an expert view on the

indicators (making use of the feedback from the expert group consultation) in which we presented information on the availability of data, the perceived reliability of the indicators, and the frequency of their use in existing rankings.

Figure 3-1: Process of Indicator Selection



The stakeholders' consultation process led to the selection of a set of indicators based on the criterion of *relevance* (according to stakeholders' perspectives). In addition, we applied four additional criteria to produce an indicators list that could be 'pre-tested':

- *Validity* – The indicator measures what it claims to measure and is not confounded by other factors. This criterion is broken down into:

- *Concept* and *construct validity*: the indicator focuses on the *performance* of (programs in) higher education and research institutions and is defined in such a way that it measures 'relative' characteristics (e.g. controlling for size of the institution).
- *Face validity*: The indicator is used in other benchmarking and/or ranking exercises and thus may be regarded as a measure of performance, which already appears to be used.

- *Reliability*: The measurement of the indicator is the same regardless of who collects the data or when the measure is repeated. The data sources and the data to build the indicator are reliable.
- *Comparability*: The indicators allow comparisons from one situation/system/location to another; broadly similar definitions are used so that data are comparable.
- *Feasibility:* The required data to construct the indicator is either available in existing databases and/or in higher education and research institutions, or can be collected with acceptable effort.

Based on the various stakeholders' and experts' assessments of the indicators as well as on our analyses using the four additional criteria, the indicators selected for the pre-test phase in U-Multirank (see 6.2) then were grouped into three categories: A, B and C, with the 'A indicators' scoring well on most criteria; the 'B-indicators' doing less well, and the 'C indicators' scoring low on most criteria. The latter indicator category was not included in the pre-test phase. During this pre-test we asked a small selection of institutions to comment on the choice of indicators, the feasibility of the data collection instruments (i.e. the questionnaires used to collect the data) as well as the clarity of the definitions for the required data elements. The intention of the pre-test was to assess the appropriateness of the conceptual and methodological instruments to be used in the (larger) pilot test.

The outcome of the pre-test was then used as further input for the wider pilot where the actual data was collected to quantify the indicators for U-Multirank at both the institutional and the field level. Based on this pilot test, the final selection of indicators was made.

3.3 Overview of indicators

Following our conceptual framework (see 2.3), the five subsections that follow present the indicators for the five dimensions (teaching & learning, research, knowledge transfer, international orientation, regional engagement). For each indicator we add a number of comments that relate to the criteria (relevance, validity, reliability, comparability, feasibility) used for the selection of the indicator.

3.3.1 Teaching and learning

Education is the core activity in most higher education and research institutions. As we noted in chapter 2, education comprises all processes to transmit knowledge, skills and values to learners (colloquially: students). Education can be conceived as a process subdivided in *enablers* (inputs,[6] process[7]) and *performance* (outputs and outcomes[8]). Teaching and learning ideally lead to the *impacts* or *benefits* that graduates will need for a successful career in the area studied and a successful, happy life as an involved citizen of a civil society. Career and quality of life are complex concepts, involving lifelong impacts. Moreover, the pace of change of higher education and research institutions means that long-term performance is of low predictive value for judgments on the future of those institutions. All we could aspire to in a ranking is to assess 'early warning indicators' of higher education's contribution, i.e. outcomes and outputs. Students' learning outcomes after graduation would be a good measure of outcomes. However, measures of *learning outcomes* that are internationally comparable are only now being developed in the AHELO project (see chapter 1)[9]. At this moment such measures do not exist, but if the AHELO project succeeds they would be a perfect complementary element in our indicator set.

Therefore, a combination of indicators was sought in order to reflect performance in the teaching and learning dimension. Teaching & learning can be looked at from different levels and different perspectives. As one of the main objectives of our U-Multirank project is to inform stakeholders such as students, their perspective is important too. From their point of view, the output to be judged is the educational process, so especially for the field-based rankings we will consider indicators that from a macro perspective are perceived as enablers.

---

[6] Inputs include resources for the education process: staff quality and quantity, facilities like libraries, books, ICT, perhaps living and sports, funding available for those resources, and student quality and quantity.

[7] The process of education includes design and implementation of curricula, with formal teaching, self study, peer learning, counselling services, etc.

[8] Outputs are direct products of a process, outcomes relate to achievements due to the outputs.

[9] http://www.oecd.org/document/22/0,3343,en_2649_35961291_40624662_1_1_1_1,00.html.

Another approach to get close to learning outcomes lies in assessing the quality of study programs. Quality assurance procedures, even if they have become almost ubiquitous in this world's higher education, are too diverse to lead to comparable indicators (see chapter 1): some quality assurance procedures focus on programs, others on entire higher education institutions; they have different foci, use different data, different performance indicators and different 'algorithms' to arrive at judgments. The qualifications frameworks currently being developed in the Bologna Process and in the EU may come to play a harmonising role with regard to educational standards in Europe, but they are not yet effective (Westerheijden et al., 2010) and of course they do not apply in the rest of the world.

Indicators of the type of studies offered have been taken into consideration as objective bases for different qualities of programs, such as their interdisciplinary character. Besides, measures of students' progressing through their programs can be seen as indicators for the quality of their learning.

Proceeding from the adage that 'quality is in the eye of the beholder', indicators for quality can be sought in student and graduate assessments of their learning experience. The student/graduate experience of education is conceptually closer to what those same students learn than judgments by external agents could be. Students' opinions may derive from investment or from consumption motives, but it is an axiom of economic theories as well as of civil society that persons know their own interest (and experience) best. Therefore we have chosen indicators reflecting both.

An issue might be whether student satisfaction surveys are prone to manipulation: do students voice their loyalty to the institution rather than their genuine (dis-)satisfaction? This is not seen as a major problem as studies show that loyalty depends on satisfaction (Athiyaman, 1997; Brown & Mazzarol, 2009; OECD, 2003). Nevertheless we should remain vigilant to uncover signs of university efforts to manipulate their students' responses; in our experience, including control questions in the survey on how and with which additional information students were approached to participate gives a good indication. Non-plausible student responses (for instance an extremely short time to complete the online questionnaire) could be eliminated.

Another issue about using surveys in international comparative studies concerns differences in culture that affect tendencies to respond in certain ways. Evidence from CHE rankings and from European surveys (e.g. EuroStudent[10]) shows, however, that student surveys can give valid and reliable information in a European context. One of the questions that we will return to later on in this report is whether a student survey about

[10] http://www.eurostudent.eu:8080/index_html.

their own program/institution can produce valid and reliable information on a global scale.

The table below lists the Teaching & Learning indicators that were selected for the pilot test of U-Multirank. The column on the right-hand side includes some of the comments and findings that came out during the stakeholder/expert consultations and the pre-testing phases of the selection process (Table 3-1).

Table 3-1: Indicators for the dimension Teaching & Learning in the Focused Institutional and Field-based Rankings

	Focused Institutional Ranking	Definition	Comments
1	**Expenditure on teaching**	Expenditure on teaching activities (including expenditure on teaching related overhead) as a percentage of total expenditure	Data available. Indicator is input indicator. Stakeholders questioned relevance.
2	**Graduation rate**	The percentage of a cohort that graduated x years after entering the program (x is the normal ('stipulated') time expected for completing all requirements for the degree times 1.5)	Graduation rate regarded by stakeholders as most relevant indicator. Shows effectiveness of schooling process. More selective institutions score better compared to (institutions in) open access settings. Sensitive to discipline mix in institution and sensitive to economic circumstances.
3	**Interdisciplinarity of programs**	The number of degree programs involving at least two traditional disciplines as a percentage of the total number of degree programs	Based on objective statistics. Relevant indicator according to stakeholders: shows teaching leads to broadly-educated graduates. But sensitive to regulatory (accreditation) and disciplinary context. Data collection and availability problematic.
4	**Relative rate of graduate (un)employment**	The rate of unemployment of graduates 18 months after graduation as a percentage of the national rate of unemployment of graduates 18 months after graduation) (for bachelor graduates and master graduates)	Reflects extent to which institution is 'in sync' with environment. Sensitive to discipline mix in institution and sensitive to (regional) economic circumstances. Data availability poses problem.

5	**Time to degree**	Average time to degree as a percentage of the official length of the program (bachelor and master)	Reflects effectiveness of teaching process. Availability of data may be a problem. Depends on the kind of programs.
	Field-based Ranking	**Definition**	**Comments**
6	**Student-staff ratio**	The number of students per fte academic staff	Fairly generally available. Is an input indicator. Depends on educational approaches. Sensitive to definitions of 'staff' and to discipline mix in institution.
7	**Graduation rate**	The percentage of a cohort that graduated after x years after entering the program (x is the normal ('stipulated') time expected for completing all requirements for the degree times 1.5)	See above institutional ranking
8	**Investment in laboratories [for Engineering FBR]**	Investment in laboratories (average over last five years, in millions in national currencies) per student	High standard laboratories essential for offering high quality education. International comparisons difficult.
9	**Qualification of academic staff**	The number of academic staff with PhD as a percentage of total number of academic staff (headcount)	Proxy for teaching staff quality. Generally available. Input indicator. Depends on national regulations and definitions of 'staff'
10	**Relative rate of graduate (un)employment**	The rate of unemployment of graduates 18 months after graduation as a percentage of the national rate of unemployment of graduates 18 months after graduation) (for bachelor graduates and master graduates)	See above institutional ranking
11	**Interdisciplinarity of programs**	The number of degree programs involving at least two traditional disciplines as a percentage of the total number of degree programs	See above institutional ranking
12	**Inclusion of issues relevant for employability in curricula**	Rating existence of inclusion into curriculum (minimum levels/standards) of: project based learning; joint courses/projects with business students (engineering); business knowledge (engineering); project management; presentation skills; existence of external advisory board (including employers)	Problems with regard to availability of data.

13	**Inclusion of work experience into the program**	Rating based on duration (weeks/credits) and modality (compulsory or recommended)	Data easily available.
14	**Computer Facilities: internet access**	Index including: hardware; internet access, including WLAN; (field specific) software; access to computer support	Data easily available.
15	**Student gender balance**	Number of female students as a percentage of total enrolment	Indicates social equity (a balanced situation is considered preferable). Generally available. But indicator of social context, not of educational quality.
	Student satisfaction indicators	**Indicators reflecting students' appreciation of several items related to the teaching & learning process.**	Student satisfaction is of high conceptual validity. It can be made available in a comparative manner through a survey. An issue might be whether student satisfaction surveys are prone to manipulation: do students voice their loyalty to the institution rather than their genuine (dis-)satisfaction? Global comparability problematic: Cross-cultural differences may affect the students' answers to the questions.
16	**Student satisfaction: Overall judgment of program**	Overall satisfaction of students with their program and the situation at their higher education institution	Refers to single question to give an 'overall' assessment; no composite indicator.
17	**Student satisfaction: research orientation of educational program**	Index of four items: research orientation of the courses, teaching of relevant research methods, opportunities for early participation in research and stimulation to give conference papers.	
18	**Student satisfaction: Evaluation of teaching**	Satisfaction with regard to student's role in the evaluation of teaching, including prevalence of course evaluation by students, relevance of issues included in course evaluation, information about evaluation outcomes, impact of evaluations	

19	**Student satisfaction: Facilities**	The satisfaction of students with respect to facilities, including: • Classrooms/lecture halls: Index including: Availability/access for students; number of places; technical facilities/devices; • Laboratories: Index including: Availability/access for students; number of places; technical facilities/devices; • Libraries: Index including: availability of literature needed; access to electronic journals; support services/e-services.	
20	**Student satisfaction: Organization of program**	The satisfaction of students with the organization of a program, including possibility to graduate in time, access to classes/courses, class size, relation of examination requirements to teaching	
21	**Student satisfaction: Promotion of employability (inclusion of work experience)**	Index of several items: Students assess the support during their internships, the organization, preparation and evaluation of internships, the links with the theoretical phases	
22	**Student satisfaction: Quality of courses**	Index including: Range of courses offered, coherence of modules/courses, didactic competencies of staff, stimulation by teaching, quality of learning materials, quality of laboratory courses (engineering)	
23	**Student satisfaction: Social climate**	Index including: • Interaction with other students • Interaction with teachers • Attitude towards students in city • Security	
24	**Student satisfaction: Support by teachers**	Included items: Availability of teachers/professors (e.g. during office hours, via email); informal advice and coaching; feedback on homework, assignments, examinations; coaching during laboratory/IT tutorials (engineering only); support during individual study time (e.g. through learning platforms); suitability of handouts.	
25	**Student satisfaction: Opportunities for a stay abroad**	Index made up of several items: The attractiveness of the university's exchange programs and the partner universities; availability of exchange places; support and guidance in preparing for stay abroad; financial	

		support (scholarships, exemption from study fees); transfer of credits from exchange university; integration of the stay abroad into studies (no time loss caused by stay abroad) and support in finding internships abroad.	
26	**Student satisfaction: Student services**	Quality of a range of student services including: general student information, accommodation services, , financial services, career service, international office and student organizations/associations	
27	**Student Satisfaction: University webpage**	Quality of information for students on the website. Index of several items including general information on institution and admissions, information about the program, information about classes/lectures; English-language information (for international students in non-English speaking countries)	

One indicator dropped from the list during the stakeholder consultation is *graduate earnings*. Although the indicator may reflect the extent to which employers value the institution's graduates, it was felt that this indicator is very sensitive to economic circumstances and institutions have little influence on labor markets. In addition, data availability proved unsatisfactory for this indicator and comparability issues negatively affect its reliability.

For our field-based rankings, subject-level approaches to quality and educational standards do exist. In business studies, the 'triple crown' of specialized, voluntary accreditation by AACSB (USA), AMBA (UK) and EQUIS (Europe) creates a build-up of expectations on study programs in the field. In the field of engineering, the Washington Accord is an 'international agreement among bodies responsible for accrediting engineering degree programs. It recognizes the substantial equivalency of programs accredited by those bodies and recommends that graduates of programs accredited by any of the signatory bodies be recognized by the other bodies as having met the academic requirements for entry to the practice of engineering' (www.washingtonaccord.org).

In general, information on whether programs have acquired one or more of these international accreditations presents an overall, distant proxy to their educational quality. However, the freedom to opt for international accreditation in business studies may differ across countries, which makes an accreditation indicator less suitable for international comparative ranking. In engineering, adherence to the Washington Accord depends on national-level agencies, not on individual higher education institutions'

strategies. These considerations have contributed to our decision not to include accreditation-related indicators in our list of Teaching & Learning performance indicators.

Instead, the quality of the learning experience is reflected in the student satisfaction indicators included in Table 3-1. These indicators can be based on a student survey carried out among a sample of students from Business Studies and Engineering. As shown in the bottom half of Table 3-1, this survey focuses on provision of courses, organization of programs and examinations, interaction with teachers, facilities, etc. Stakeholders' feedback on the student satisfaction indicators revealed that they have a positive view overall of the relevance of the indicators on student satisfaction. However, it was also felt that the total number of indicators is quite high and should be reduced in the final indicator set.

In the field-based rankings, objective indicators are used in addition to the student satisfaction indicators. Most are similar to the indicators in the focused institutional rankings. Some additional indicators are included to pay attention to the facilities and services provided by the institution to enhance the learning experience (e.g. laboratories, curriculum).

3.3.2 Research

Selecting indicators for capturing the research performance of a higher education and research institution or a disciplinary unit (e.g. department, faculty) within that institution has to start with the definition of *research*. We take the definition set out in OECD's *Frascati Manual*:[11]

Research and experimental development (R&D) comprise creative work undertaken on a systematic basis in order to increase the stock of knowledge, including knowledge of man, culture and society, and the use of this stock of knowledge to devise new applications.

The term R&D covers three activities: basic research, applied research and experimental development. Given the increasing complexity of the research function of higher education institutions and its extension beyond PhD awarding institutions, U-Multirank adopts a broad definition of research, incorporating elements of both basic and practice-oriented (applied) research. There is a growing diversity of research missions across the classical research universities and the more vocational oriented institutions (university colleges, institutes of technology, universities of applied sciences, Fachhochschulen, etc). This is reflected in the wide range of research outputs and outlets mapped across the full spectrum, from discovery to knowledge transfer to innovation.

[11] http://browse.oecdbookshop.org/oecd/pdfs/browseit/9202081E.PDF

Research performance indicators may be distinguished into:

- *Output indicators,* measuring the quantity of research products. Typical examples are the number of papers published or the number of PhDs delivered.

- *Outcome indicators,* relating to a level of performance or achievement. For instance the contribution research makes to the advancement of scientific scholarly knowledge. Typical examples are citation rates, awards and prizes.

- *Impact indicators,* referring to the contribution of research outcomes to society, culture, the environment and/or the economy.

Given that in most disciplines publications are often seen as the single most important research output of higher education institutions, research performance measurement frequently takes place through bibliometric data. Data on publications, texts and citations is readily available for building bibliometric indicators (see Table 3-2). This is much less the case for data on research awards and data underlying impact indicators. In addition to performance measures, sometimes input-related proxies such as the volume of research staff and research income are in use to describe the research taking place in a particular institution or unit. Compared to such input indicators, bibliometric indicators may be more valid measures for the output or productivity of research teams and institutions. Increasingly sophisticated indicators such as citation indexes and co-citation indicators have been developed over time. However, an important issue in the production of bibliometric indicators lies in the definition of items that are considered as relevant.

The Expert Group on Assessment of University Based Research[12] defines research output as referring to individual journal articles, conference publications, book chapters, artistic performances, films, etc. While journals are the primary publication channel for almost all disciplines, their importance differs across disciplines. In some fields, books (monographs) play a major role, while book chapters or conference proceedings have a higher status in other fields (see Table 3-2). Therefore, focusing only on journal articles may not do justice to the research performance in particular disciplines. Moreover, the complexity of knowledge has led to a diverse range of output formats and research outlets. One may mention audio visual recordings, computer software and databases, technical drawings, designs or working models, major works in production or exhibition and/or award-winning design, patents or plant breeding rights, major art works, policy documents or briefs, research or technical reports, legal cases, maps, translations or editing of major works within academic standards.

[12] See: http://www.kowi.de/Portaldata/2/Resources/fp/assessing-europe-university-based-research.pdf

Table 3-2: Primary form of written communications by discipline group

	Natural sciences	Life sciences	Engineering Sciences	Social sciences & Humanities	Arts
Journal article	X	X	X	X	X
Conference proceedings			X		
Book chapters				X	
Monographs/Books				X	
Artefacts					X
Prototypes			X		

Source: Expert Group on Assessment of University-Based Research (2010)

Apart from using existing bibliometric databases , there is also the option to ask institutions themselves to list their research products, without restrictions on the type, medium or quality. While this may improve data coverage, such self-reported accounts may not be standardized or reliable, because respondents may interpret the definitions differently. For example, they may overestimate unpublished but accepted articles. This means that in the case of field-based rankings, the choice of one of these options will depend on the field.

The indicators for research performance in the focused institutional rankings and the field-based rankings are listed below (Table 3-3), along with some comments reflecting their assessment (by stakeholders and experts) against the criteria discussed in the first section of this chapter. The indicators in the table are used in the pilot test (chapters 5 and 6). The majority of the indicators are normalized by taking into account measures of an institution's (or a department's) size – that is: referring to total staff (in fte or headcounts), total revenues or other volume measures.

Table 3-3: Indicators for the dimension Research in the Focused Institutional and Field-based Rankings

	Focused Institutional Ranking	Definition	Comments
1	**Expenditure on research**	The amount of money spent on research activities in the reference year as a percentage of total expenditure	Reflects involvement in (and priority attached to) research. Thus *input* indicator. Data mostly available. Recommended by Expert Group on University-based Research. Difficult to separate teaching and research expenditure in a uniform way.
2	**Research income from competitive sources**	Income from European research programs + income from other international competitive research programs + income from research councils + income from privately funded research contracts as a share of total income	Success in winning grants indicates quality of research. Expert Group regards the indicator as relevant. Levels of external funding may vary greatly across disciplines and countries. Lack of clear delineation affects comparability. In some countries, competitive public funding may be difficult to separate from other public funding.
3	**Research publication output**	Frequency count of research publications with at least one author address referring to selected institution (within Web of Science)	Broadly accepted. Data largely available. Widely used in research rankings (Shanghai, Leiden ranking, HEEACT). Different disciplinary customs cause distortion. Since publications are in peer-reviewed journals, they also signify a certain degree of research quality. However, focus on peer reviewed journal articles is too narrow for some disciplines.
4	**Post-doc positions (share)**	Number of post-doc positions /fte academic staff	Success in attracting post-docs indicates quality of research. Reliability affected by the contextual characteristics of a country's science system. Definitions may vary across countries. Data availability may be weak.

5	**Interdisciplinary research activities**	Share of research publications authored by multiple units from the same institution (based on self-reported data)	Research activities are increasingly becoming interdisciplinary. Indicator may be difficult to define (and collect) satisfactory.
6	**Field-normalized citation rate**	Field-normalized citation impact score, where the fields are equivalent to the Thomson Reuters Journal Categories. 'Actual' citation counts are compared to 'expected' counts based on the average impact score of all journals assigned to a field. A score larger than one represents a citation impact above world average within that field of science, whereas scores below one represent below average impact.	Indicates international scientific impact. Widely used and accepted indicator, especially in the exact sciences. Certain parts of social sciences, humanities and engineering are less well covered by citation indexes. Disregards impact of publications in journals aimed at professional audience.
7	**Share of highly cited research publications**	Share of top 10% most highly cited publications; comparing 'actual' citation counts to 'expected' counts per field; citation impact distributions are calculated by applying a fixed citation-window, for two 'research-based' document types: articles, reviews. These data refer to database years.	Publishing in top-ranked, high impact journals reflects quality of research. Indicator relevant primarily for exact/natural sciences. Data largely available. Books and proceedings are not considered. Never been used before in any international classification or ranking.
8	**Number of art related outputs**	Count of all relevant research-based tangible outputs in creative arts /fte academic staff	Recognizes outputs other than publications (e.g. exhibition catalogues, musical compositions, designs). This allows musical academies and art schools to be covered in ranking. Data suffers from lack of agreed definitions and lack of availability. Quantities difficult to aggregate.
9	**Number of international awards and prizes won for research work**	Prizes, medals, awards and scholarships won by employees for research work and in (inter-) national cultural competitions, including awards granted by academies of science.	Indicator of peer esteem. Recognition of quality. Data suffers from lack of agreed definitions and lack of availability. Quantities difficult to aggregate. Comparison across disciplines difficult.

	Field-based Ranking	Definition	Comments
10	**External research income**	Level of funding attracted by researchers from contracts with external sources, including competitive grants and research income from government, industry, business and community organizations, as a percentage of total income	Success in winning grants indicates quality of research. Lack of clear delineation affects comparability. Annual and accurate numbers hard to retrieve, research contracts may run over several years.
11	**Research publication output**	Frequency count of (Web of Science) research publications with at least one author address referring to selected institutional unit (relative to fte academic staff)	Frequently used indicator. However, research findings are not just published in journals.
12	**Doctorate productivity**	Number of completed PhDs per number of Professors (head count)*100 (three-year average)	Indicates aspects of the quantity and quality of a unit's research. Indicator affected by the contextuality of a country's science system.
13	**Field normalized citation rate**	See definition under Institutional Ranking	See comments made above for corresponding entry under Institutional Ranking
14	**Highly cited research publications**	See definition under Institutional Ranking	Top-end citation indices are less useful in some fields where high-profile research findings are also published in other outlets (books, reports, conference proceedings).

Bibliometric indicators (citations, publications) are part of every research-based ranking. To acknowledge the output in the arts, an indicator reflecting arts-related output is included in U-Multirank as well. However, data availability is posing some challenges here. Research publications other than peer-reviewed journal publications are included, but this requires self-reporting by institutions based on clear definitions of the types of publications.

An indicator that was considered for use but dropped is 'Presence of research related promotion schemes for academic staff'. A performance-based appraisal/ incentive system (e.g. tenure track system) may increase the attractiveness of an institution to strong researchers, but it proved difficult to define such an indicator in a uniform way across multiple contexts (institutions, borders, disciplines).

Yet another indicator excluded during the process is 'Share of within-country joint research publications'. The number of publications that involve at least one author from

another organization in the same country reflects successful national research cooperation. While such data is available, it is limited only to national authors. During the indicator selection process the relevance of the indicator was questioned, more so given the fact that research often is an international endeavor.

Some of the indicators in Table 3-3 are of an *input*-type, such as expenditure on research, competitive grants and post-doc positions. However, stakeholders regarded them as relevant, even though data availability and definitions may sometimes pose a challenge. Therefore it was decided to keep them in the list of indicators for U-Multirank's institutional ranking.

Indicators for reflecting research performance in the field-based rankings are fewer in number. The ones that are included are largely overlapping with indicators for the institutional ranking. The fact that they are relating to a particular field opens up the door for additional indicators, i.e. doctoral productivity.

After pre-testing the indicators it has become clear that there are some data availability issues – in terms of a clarity of definitions (for instance FTE staff) and the cost of collecting particular indicators. The pre-test also revealed that there may be cases where numbers (e.g. art-related outputs) need to be estimated by the reporting institutions and departments. This may affect reliability. A test of the indicators (and the underlying data elements) in the more broad pilot study (see chapters 5 and 6), however, allows us to come to a firmer conclusion on the final list of indicators for the dimension of research.

3.3.3 Knowledge transfer

Knowledge transfer has become increasingly relevant for higher education and research institutions as many nations and regions strive to make more science output readily available for economic, social and cultural development. There are large differences between efforts and performance of individual institutions in this respect, partly because of the official mandate of an institution and partly because of the strategic profile chosen by individual institutions. *Knowledge transfer* is a broader and more encompassing concept than *technology transfer*. It may be defined as:

> The process by which the knowledge, expertise and intellectually linked assets of Higher Education Institutions are constructively applied beyond Higher Education for the wider benefit of the economy and society, through two-way engagement with business, the public sector, cultural and community partners. (Holi et al., 2008).

Measuring the impact of the knowledge transfer (or: the knowledge exchange) process in higher education and research institutions and ultimately on users, i.e. business and the economy, has now become a preoccupation of many governing and funding bodies, as well as policy-makers. So far, most attention has been devoted to measuring Technology

Transfer (TT) activities. Traditionally TT is primarily concerned with the management of intellectual property (IP) produced by universities and other higher education and research institutions. TT means identifying, protecting, exploiting and defending intellectual property (OECD, 2003). Higher education and research institutions often have *technology transfer offices* (TTOs) (Debackere & Veugelers, 2005), which are units that liaise with industry and assist higher education and research institutions' personnel in the *commercialisation* of research results. TTOs provide services in terms of assessing inventions, patenting, licensing IP, developing and funding spin-offs and other start-ups and approaching firms for contract based arrangements.

The broader nature of Knowledge Transfer compared to TT also means it includes other forms – *channels* – of transfer than those requiring strong IP protection. A typical classification of mechanisms and channels for knowledge transfer between higher education and research institutions and other actors would include four main interaction channels for communication between higher education and research institutions and their environment:

- Texts, including scientific, professional and popular,
- People, including students and researchers,
- Artefacts, including equipment, protocols, rules and regulations,
- Money.

Texts are an obvious knowledge transfer channel. Publishing in scientific or popular media is, however, already covered under the research dimension in U-Multirank. In the case of texts, it is customary to distinguish between two forms: *publications*, where copyright protects how ideas are expressed but not the ideas themselves, and *patents*, which grant exclusive rights to use the inventions explained in them. While publications are part of the research dimension in U-Multirank, patents will be included under the Knowledge Transfer dimension.

People is another channel of knowledge transfer. People carry with them competences, skills and *tacit* knowledge. Indeed, many knowledge exchanges will be person-embodied. This type of knowledge transfer, however, is captured through the Teaching & Learning and Regional Orientation dimensions included in U-Multirank. Knowledge transfer through people also takes place through networks, continuous professional development (CPD)[13] and research contracts.

[13] CPD may be defined as: The means by which members of professional associations maintain, improve and broaden their knowledge and skills and develop the personal qualities required in their professional lives, usually through a range of short and long training programmes (offered by education institutions), some of which have an option of accreditation.

Money flows are an important interaction channel, next to texts and people. Unlike texts and people, money is not a carrier of knowledge, but a way of valuing the knowledge transferred in its different forms. The money involved in contract research, CPD, consultancy and commercialisation is one of the traditional indicators of knowledge exchange, often used in surveys of TTOs, such as the one carried out by the US-based Association of University Technology Managers (AUTM) for its Annual Licensing survey.

Artefacts make up the fourth major channel of interaction. Artefacts are concrete, physical forms in which knowledge can be carried and transferred. They are more or less 'ready to use', such as machinery, software, new materials or modified organisms. This is often called 'technology'. Artefacts may also extend to art-related outputs produced by scholars working in the arts and humanities disciplines. These works of art, including artistic performances, films and exhibition catalogues have been included in the scholarly outputs covered in the Research dimension of U-Multirank.

Most approaches to knowledge transfer measurement primarily address revenues obtained from the commercialization of Intellectual Property (IP). Clearly the measurement of income from IP is an incomplete reflection of knowledge transfer performance. For this reason, new approaches have been developed, such as the Higher Education-Business and Community Interaction (HE-BCI) Survey in the UK. [14] This UK survey began in 2001 and recognises a broad spectrum of activities with both financial and non-financial objectives. However, it remains a fact that many indicators in the area of Knowledge Transfer are still in their infancy — in particular the ones that try to go beyond the IP issues.[15] Moreover, there is a need to define knowledge transfer more clearly in order to delineate it from dimensions such as Teaching, Research and Regional Engagement. Like research, knowledge transfer is a process, where inputs, throughputs, outputs and outcomes may be distinguished. Most knowledge transfer measurements focus on the input, some on the output and even fewer on the outcome (or impact) side of this process.

U-Multirank particularly wants to capture aspects of knowledge transfer *performance*. However, given the state of the art in measuring knowledge transfer (Holi et al., 2008)

[14] http://ec.europa.eu/invest-in-research/pdf/download_en/knowledge_transfer_web.pdf. The HE-BCI survey is managed by the Higher Education Funding Council for England (HEFCE) and used as a source of information to inform the funding allocations to reward the UK universities' third stream activities. See: http://www.hefce.ac.uk/econsoc/buscom/hebci/

[15] The European Commission-sponsored project E3M (Montesinos et. al., 2008) aims to create a ranking methodology for measuring university third mission activities along three subdimensions: Continuing Education (CE), Technology Transfer & Innovation (TT&I) and Social Engagement (SE).

and the near absence of (internationally comparable) data (see chapter 4)[16], it proved extremely difficult to do so. Most candidates for additional indicators are of an input-type.

The knowledge transfer indicators are presented in Table 3-4, together with – in the right hand column – some of the pros and cons of the indicators expressed by experts and stakeholders during the indicator selection process. The first selection of indicators was inspired by the international literature on knowledge transfer metrics and existing surveys in this area. An important reference is the report published in 2009 by the Expert Group on Knowledge Transfer Metrics (EGKTM) set up by DG Research of the European Commission.[17]

Table 3-4: Indicators for the dimension Knowledge Transfer (KT) in the Focused Institutional and Field-based Rankings

	Focused Institutional Ranking	Definition	Comments
1	**Incentives for Knowledge Exchange**	Presence of knowledge exchange activities as part of the performance appraisal system	Such a scheme encourages staff to engage in KT. Information available in institutions. Difficult to define uniformly across institutions, borders, disciplines. New indicator.
2	**Third Party Funding**	The amount of income for cooperative projects that are part of public programs (e.g. EC Framework programs) plus direct industry income as a proportion of total income	Signals KT success. Some data do exist (although definitions may vary). Is regarded as relevant indicator by EGKTM.
3	**University-industry joint publications**	Relative number of research publications that list an author affiliate address referring to a business enterprise or a private sector R&D unit; relative to fte academic staff	Indicates appreciation of research by industry. Reflects successful partnerships. Less relevant for HEIs oriented to humanities, social sciences. ISI databases available. Used in CWTS University-Industry Research Cooperation Scoreboard.

[16] See also the brief section on the EUMIDA project, included in this report. One of EUMIDA's findings is that data on technology transfer activity and patenting is difficult to collect in a standardized way (using uniform definitions, etc.)

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

4	**Patents**	The number of patent applications for which the university acts as an applicant related to number of academic staff	Widely used in KT surveys. Included in U-Map. Depends on disciplinary mix of HEI. Data are available from secondary (identical) data sources.
5	**Size of Technology Transfer Office**	Number of employees (FTE) at Technology Transfer Office related to the number of FTE academic staff	Reflects priority for KT. Input indicator, could also show inefficiency. Data are mostly directly available. KT function may be dispersed across the HEI. Not regarded as core indicator by EGKTM.
6	**CPD courses offered**	Number of CPD courses offered per academic staff (fte)	Captures outreach to professions. Relatively new indicator. CPD difficult to describe uniformly.
7	**Co-patents**	Percentage of university patents for which at least one co-applicant is a firm, as a proportion of all patents	Reflects extent to which HEI shares its IP with external partners. Not widely used in TT surveys. Depends on disciplinary mix of HEI. Data available from secondary sources (PatStat).
8	**Number of Spin-offs**	The number of spin-offs created over the last three years per academic staff (fte)	EGKTM regards Spin-offs as core indicator. Data available from secondary sources. Clear definition and demarcation criteria needed. Does not reveal market value of spin-offs.
	Field-based Ranking	**Definition**	**Comments**
9	**Academic staff with work experience outside higher education**	Percentage of academic staff with work experience outside higher education within the last 10 years	Signals that HEI's staff is well-placed to bring work experience into their academic work. Data difficult to collect.
10	**Annual income from licensing**	The annual income from licensing agreements as a percentage of total income	Licensing reflects exploiting of IP. Indicator is used widely. HEIs not doing research in natural sciences/engineering/ medical sciences hardly covered.
11	**Co-patents**	Percentage of university patents for which at least one co-applicant is a firm, as a proportion of all patents	See above institutional ranking

12	**Joint research contracts with private sector**	Budget and number of joint research projects with private enterprises per fte academic staff	Indicator of (applied) R&D activities. Indicator only refers to the size of projects, not their impact in terms of KT.
13	**Number of license agreements**	The number of licence agreements as a percentage of the number of patents	Licensing reflects exploiting of IP. Indicator is used widely. HEIs not doing research in natural sciences/engineering/ medical sciences hardly covered. Number of licences more robust than licensing income.
14	**Patents awarded**	The number of patents awarded to the university related to number of academic staff	Widely used KT indicator. Data available from secondary (identical) data sources. Patents with an academic inventor but another institutional applicant(s) not taken into account. Not relevant for all fields.
15	**University-industry joint publications**	Number of research publications that list an author affiliate address referring to a business enterprise or a private sector R&D unit, relative to fte academic staff	See above institutional ranking. Differences in relevance by fields.

Cultural awards and prizes won in (inter)national cultural competitions would be an additional indicator that goes beyond the traditional technology-oriented indicators. However, the indicator is difficult to define uniformly and from the pre-test it became clear that data is difficult to collect. Therefore this indicator was not kept in the list for the pilot.

While there is a large overlap in terms of indicators between the institutional ranking and the field-based ranking, the indicators related to licensing were felt to be less relevant for the institution as a whole. Licensing income is part of the third party funding indicator for the institutional level though. The number of collaborative research projects (university-industry) is another example of a knowledge transfer indicator that was not selected for the Focused Institutional Ranking.

3.3.4 International orientation

Internationalization is a widely discussed and complex phenomenon in higher education. The rise of globalization and Europeanization have put growing pressure on higher education and research institutions to respond to these trends and develop an international orientation in their activities. Internationalization activities can be categorized in three types (Teichler, 2004):

- Activities to develop and promote international mobility of students and staff,
- Activities to develop and enhance international cooperation,
- Activities to develop and increase international competition.

The rationales that drive these activities are diverse. Among others, they comprise (IAU, 2005):

- The increasing emphasis on the need to prepare students international labor markets and to increase their international cultural awareness,
- The increasing internationalization of curricula
- The wish to increase the international position and reputation of higher education and research institutions (Enquist, 2005).

In the literature (Brandenburg, 2007; Enquist, 2005; Nuffic, 2010; IAU, 2005) many indicators have been identified, most of which refer to inputs and processes. The outcomes and impacts of internationalization activities are not very well covered by existing internationalization indicators.

For many of the indicators data are available in the institutional databases. Hardly any of such data can be found in national or international databases.

The various manifestations and results of internationalization are captured through the list of indicators shown in Table 3-5. The table includes some comments made during the consultation process that led to the selection of the indicators.

Table 3-5: Indicators for the dimension International Orientation in the Focused Institutional and Field-based Rankings

	Focused Institutional Ranking	Definition	Comments
1	**Educational programs in foreign language**	The number of programs offered in a foreign language as a percentage of the total number of programs offered	Signals the commitment to international orientation in teaching and learning. Data availability good. Relevant indicator. Used quite frequently. Sensitive to relative 'size' of national
2	**International academic staff**	Foreign academic staff members (headcount) as percentage of total number of academic staff members (headcount). Foreign academic staff is academic staff with a foreign	Considered to be relevant by stakeholders. Nationality not the most precise way of measuring international orientation.

		nationality, employed by the institution or working on an exchange base	
3	**International doctorate graduation rate**	The number of doctorate degrees awarded to students with a foreign nationality, as a percentage of the total number of doctorate degrees awarded	Indicator not used frequently. Some stakeholders see it as less relevant. Availability of data problematic.
4	**International joint research publications**	Relative number of research publications that list one or more author affiliate addresses in another country relative to research staff	Only indicator addressing research internationalization. Data available in international data bases, but bias towards certain disciplines and languages.
5	**Number of joint degree programs**	The number of students in joint degree programs with foreign university (including integrated period at foreign university) as a percentage of total enrolment	Integration of international learning experiences is central element of internationalization. Data available. Indicator not often used.
	Field-based Ranking	**Definition**	**Comments**
6	**Incoming and outgoing students**	Incoming exchange students as a percentage of total number of students and the number of students going abroad as a percentage of total number of students enrolled	Important indicator of the international 'atmosphere' of a faculty/department. Addresses student mobility and curriculum quality. Data available.
7	**International graduate employment rate**	The number of graduates employed abroad or in an international organization as a percentage of the total number of graduates employed	Indicates the student preparedness on the international labor market. Data not readily available. No clear international standards for measuring.
8	**International academic staff**	Percentage of international academic staff in total number of (regular) academic staff	See above institutional ranking
9	**International research grants**	Research grants attained from foreign and international funding bodies as a percentage of total income	Proxy of the international reputation and quality of research activities. Data are available. Stakeholders question relevance.
10	**Student satisfaction: Internationalization of programs**	Index including the attractiveness of the university's exchange programs, the attractiveness of the partner universities, the sufficiency of the number of exchange places; support and guidance in preparing the stay abroad; financial support; the transfer of credits from exchange university; the integration of	Addresses quality of the curriculum. Not used frequently.

		the stay abroad into studies (no time loss caused by stay abroad).	
11	**Joint international publications**	Relative number of research publications that list one or more author affiliate addresses in another country relative to academic staff	See above institutional ranking, but no problems of disciplinary distortion because comparison is made within the field.
12	**Percentage of international students**	The number of degree-seeking students with a foreign diploma on entrance as percentage of total enrolment in degree programs.	Reflects attractiveness to international students. Data available but sensitive to location (distance to border) of HEI. Stakeholders consider the indicator important.
13	**Student satisfaction: International orientation of programs**	Rating including several issues: existence of joint degree programs, inclusion of mandatory stays abroad, international students (degree and exchange), international background of staff and teaching in foreign languages.	Good indicator of international orientation of teaching; composite indicators depend on the availability of each data element.

It should be pointed out here that one of the indicators is a *student satisfaction indicator*: 'Student satisfaction: Internationalisation of programs'. This describes the opportunities for students to go abroad. Students' judgments about the opportunities to arrange a semester or an internship abroad are an aspect of the internationalization of programs. This indicator is relevant for the field level.

An indicator that was considered, but dropped during the stakeholders' consultation process is 'Size of international office'. While this indicates the commitment of the higher education and research institution to internationalization, and data is available, stakeholders consider this indicator not very important. Moreover, the validity is questionable as the size of the international office as a facilitating service is a very distant proxy indicator.

The indicator 'international graduate employment rate' was dropped from the list for focused institutional rankings because a large majority of stakeholders judged this to be insufficiently relevant. At the field level this indicator was however seen as an attractive indicator for the international orientation of the program.

'International partnerships', that is the number of international academic networks a higher education and research institution participates in, is a potential indicator of the international embeddedness of the institution (department). However, it was dropped from the list during the stakeholder consultation as there is no clear internationally accepted way of counting partnerships. The same argument was used to exclude the indicator 'Joint international research projects'.

3.3.5 Regional engagement

The *region* has become an important entity in the processes of economic and social development and innovation. Gaps between regions in terms of these processes are growing and regions that have skilled people and the infrastructure for innovation have a competitive advantage (Ischinger et al., 2009). Higher education and research institutions can play an important role in the process of creating the conditions for a region to prosper. Creating and expanding this role in the region has become highly relevant for many public policymakers at the national and regional level, as well as for institutional administrators. How well a higher education and research institution is engaged in the region is increasingly considered to be an important part of the mission of higher education institutions.

Regional engagement is part of the broader concept of the 'third mission' of an institution. In the European project on third mission ranking (Montesinos, 2008) this 'third mission' consists of three dimensions: a social dimension, an enterprise dimension and an innovation dimension. The latter two dimensions are covered in the U-Multirank dimension 'Knowledge Transfer'. Indicators for the social dimension of the third mission comprise indicators on international mobility (that are covered in the U-Multirank dimension International Orientation) and a very limited number of indicators on regional engagement.

Activities and indicators on regional and community engagement can be categorized in three groups: outreach, partnerships and curricular engagement[18]. Outreach focuses on the application and provision of institutional resources for regional and community use, benefitting both university and the regional community. Partnerships focus on collaborative interactions with the region/community and related scholarship for the mutual beneficial exchange, exploration, discovery and application of knowledge, information and resources. Curricular engagement refers to teaching, learning and scholarship that engage faculty, students and region/community in mutual beneficial and respectful collaboration.

Both enabling indicators and performance indicators are suggested in the literature on regional and community engagement. However, most attention is paid to the enablers and to indicators addressing the way an institution organizes its engagement activities. These indicators are based on checklists assessing the extent to which regional engagement is part of the institutional mission and integrated in the routines and procedures of the institution. Do the reward and promotion schemes of the institution

[18] See: http://classifications.carnegiefoundation.org/details/community_engagement.php

acknowledge regional engagement activities? Are there visible structures that function to assist with region-based teaching and learning? Is there adequate funding available for establishing and deepening region-based activities? Are there courses that have a regional component (such as service-learning courses)? Are there mutually beneficial, sustained partnerships with regional community partners? These are typical items on such checklists (Furco et al, 2009; Hollander et al, 2001). The problem with these checklists is that the information is not readily available. Institutional or external assessors need to collect the information, which makes the robustness and reliability of the results in an international comparative setting highly questionable.

Other indicators for regional engagement capture the relative size of the interaction. How much does the institution draw on regional resources (students, staff, funding) and how much does the region draw on the resources provided by the higher education and research institution (graduates and facilities)?

Clarification is required as to what constitutes a region. U-Multirank has suggested to start with the existing list of regions in the Nomenclature of Territorial Units for Statistics (NUTS) classification developed and used by the European Union[19], in particular the NUTS 2 level. For non-European countries a different region classification will need to be used. The idea is to make use of the lower level (Territorial level 3) of the OECD classification of its member states. This is composed of micro-regions[20]. As it is with most standard lists, they work fine in the majority of cases, but there are always cases where a different definition is more appropriate. In our feasibility study, we have allowed higher education and research institutions to specify their own delimitation of region if they feel there are valid reasons for doing so. Table 3-6 includes the indicators on regional engagement, along with the comments made during the stakeholder and expert consultations.

Table 3-6: Indicators for the dimension Regional Engagement in the Focused Institutional and Field-based Rankings

	Focused Institutional Ranking	Definition	Comments
1	**Graduates working in the region**	The number of graduates working in the region, as a percentage of all graduates employed	Frequently used in benchmarking exercises. Stakeholders like indicator. No national data on graduate destinations.

[19]http://epp.eurostat.ec.europa.eu/portal/page/portal/region_cities/regional_statistics/nuts_classification

[20] http://www.oecd.org/document/62/0,3343,en_2649_34413_36878718_1_1_1_1,00.html

2	**Income from regional/local sources**	Institutional income from local regional authorities, local/regional charities and local/regional contracts as a percentage of total institutional income	Reflects connection and engagement with regional/local society. Sensitive to way public funding for HEI is organized (national versus regional/federal systems). Availability of data problematic.
3	**Regional joint research publications**	Number of research publications that list one or more author-affiliate addresses in the same NUTS2 or NUTS3 region, relative to fte academic staff	Reflects 'local' research cooperation. Data available (Web of Science), but professional (laymen's) publications not covered.
4	**Research contracts with regional business**	The number of research projects with regional firms, as a proportion of the total number of collaborative research projects	Seen as valid and relevant indicator, hardly any records kept on (regional) contracts. New type of indicator.
5	**Student internships in local/regional enterprises**	The number of student internships in regional enterprises as a percentage of total enrolment (with defined minimum of weeks and/or credits)	Internships open up communication channels between HEI and regional/local enterprises. Stakeholders see this as important indicator. Definition of internship problematic and data not readily available. Disciplinary bias.
	Field-based Ranking	**Definition**	**Comments**
6	**Degree theses in cooperation with regional enterprises**	Number of degree theses in cooperation with regional enterprises as a percentage of total number of degree theses awarded, by level of program	Reflects regional cooperation and curricular engagement. Indicator hardly ever used.
7	**Graduates working in the region**	The number of graduates working in the region, as a percentage of all graduates employed	See above institutional ranking.
8	**Regional participation in continuing education**	Number of regional participants (coming from NUTS3 region where HEI is located) as percentage of total number of population in NUTS3 region aged 25+	Indicates how much the HEI draws on the region and vice versa. Covers important aspect of curricular engagement. Data not readily available. Indicator hardly ever used.
9	**Student internships in local/regional enterprises**	Number of internships of students in regional enterprises (as percentage of total students	See above institutional ranking, but disciplinary bias not problematic at field level.
10	**Summer school/courses for secondary education students**	Number of participants in schools/courses for secondary school students as a percentage of total enrolment	Addresses outreach activities. Limited availability of data. Lack of internationally accepted definition of summer school courses.

During the process of selection of indicators the list of indicators underwent a number of revisions. While the table shows the indicators that were placed in categories A and B – as indicators to be included in the pilot test – some other indicators were placed in category C (not to be included in the pilot phase). In the dimension Regional Engagement there are a number of such 'C indicators':

- 'Co-patents with regional firms' reflect cooperative research activities between higher education institutions and regional firms. While data may be found in international patent databases, the indicator is not often used and stakeholders did not particularly favor the indicator. Therefore it was dropped from our list.
- The same holds for measures of the regional economic impact of a higher education institution, such as the number of jobs generated by the university. Assessing what the higher education and research institution 'delivers' to the region (in economic terms) is seen as most relevant but data constraints prevent us from the use of such an indicator.
- Public lectures that are open to an external, mostly local audience, are a way to intensify contacts to the local community. However, stakeholders felt this indicator not to be relevant.
- A high percentage of new entrants from the region may be seen as the result of the high visibility of regionally active higher education and research institutions. It may also be a result of the engagement with regional secondary schools. This indicator however was not included in our list, mainly because it was not considered to be that relevant.

The above discussion makes it clear that regional engagement is a dimension that poses many problems with regard to availability of performance-oriented indicators and their underlying data. In the next chapter we will discuss the data gathering instruments that are available more extensively. In chapters 5 and 7the pilot study on the empirical feasibility assessment of the U-Multirank tool and its various indicators will be discussed. As a result of this pilot assessment the final list of indicators will be presented.

4 Constructing U-Multirank: databases and data collection tools

4.1 Introduction

In this chapter we will describe the databases and data collection instruments used in constructing U-Multirank. The first part is an overview of existing databases – mainly on bibliometrics and patents. The second presents an explanation of the questionnaires and survey tools used for collecting data from the institutions (the self-reported data) – at the institutional and department levels – and from students.

4.2 Databases

4.2.1 Existing databases

One of the activities in the U-Multirank project was to review existing rankings and explore their underlying databases. If existing databases can be relied on for quantifying the U-Multirank indicators this would be helpful in reducing the overall burden for institutions in handling the U-Multirank data requests. However, from the overview of classifications and rankings presented in chapter 1 (section 1.3) it is clear that international databases holding information at institution level or at lower aggregation levels are currently available only for particular aspects of the dimensions Research and Knowledge Transfer. For other aspects and dimensions, U-Multirank will have to rely on self-reported data. Regarding research output and impact, there are worldwide databases on journal publications and citations. For knowledge transfer, the database of patents compiled by the European Patent Office is available. In the next two subsections, available bibliometric and patent databases will be discussed.

To further assess the availability of data covering individual higher education and research institutions, the results of the EUMIDA project were also taken into account.[21] The EUMIDA project (see: www.eumida.org) seeks to develop the foundations of a coherent data infrastructure (and database) at the level of individual higher education institutions. Section 4.2.4 presents an overview of availability based on the outcomes of the EUMIDA project. Our analysis on data availability was completed with a brief online consultation with the group of international experts connected to U-Multirank (see section 4.2.5). The international experts were asked to give their assessment of the

[21] The U-Multirank project was granted access to the preliminary outcomes of the EUMIDA project in order to learn about data availability in the countries covered by EUMIDA.

situation with respect to data availability in some of the non-EU countries included in U-Multirank.

4.2.2 Bibliometric databases

There are a number of international databases which can serve as a source of information on the research output of a higher education and research institution (or one of its departments). An institution's quantity of research-based publications (per capita) reflects its research output and can also be seen as a measure of scientific merit or quality. In particular, if its publications are highly cited within the international scientific communities this may characterize an institution as high-impact and high-quality. The production of publications by a higher education and research institute not only reflects research activities in the sense of original scientific research, but usually also the presence of underlying capacity and capabilities for engaging in sustainable levels of scientific research.[22] The research profile of a higher education and research institution can be specified further by taking into account its engagement in various types of research collaboration. For this, one can look at joint research publications involving international, regional and private sector partners. The subset of jointly authored publications is a testimony of successful research cooperation.

Data on numbers and citations of research publications are covered relatively well in existing databases. Quantitative measurements and statistics based on information drawn from bibliographic records of publications are usually called 'bibliometric data'. These data concern the quantity of scientific publications by an author or organisation and the number of citations (references) these publications have received from other research publications. There is a wide range of research publications available for characterizing the research profile and research performance of an institution by means of bibliometric data: lab reports, journal articles, edited books, monographs, etc. The bibliometric methodologies applied in international comparative settings such as U-Multirank usually draw their information from publications that are released in scientific and technical journals. This part of the research literature is covered ('indexed') by a number of international databases. In most cases the journals indexed are internationally peer-reviewed, which means that they adhere to international quality standards. U-Multirank therefore makes use of international bibliometric databases to compile some of its research performance indicators and a number of research-related indicators belonging to the dimensions of Internationalisation, Knowledge Transfer and Regional Engagement.

[22] This is why research publication volume is a part of the U-Map indicators that reflect the activity profile of an institution.

Two of the most well-known databases that are available for carrying out bibliometric analyses are the *Web of Science* and *Scopus*.[23] Both are commercial databases that provide global coverage of the research literature and both are easily accessible. The Web of Science database is maintained by ISI, the Institute for Scientific Information, which was taken over by Thomson Reuters a few years ago. The Web of Science currently covers about 1 million new research papers per year, published in over 10,000 international and regional journals and book series in the natural sciences, social sciences, and arts and humanities. According to the Web of Science website, 3,000 of these journals account for about 75% of published articles and over 90% of cited articles.[24] The Web of Science claims to cover the highest impact journals worldwide, including Open Access journals and over 110,000 conference proceedings.

The Scopus database was launched in 2004 by the publishing house Elsevier. It claims to be the largest abstract and citation database containing both peer-reviewed research literature and web sources. It contains bibliometric information covering some 17,500 peer-reviewed journals (including 1,800 Open Access journals) from more than 5,000 international publishers. Moreover it holds information from 400 trade publications and 300 book series, as well as data about conference papers from proceedings and journals.

To compile the publications-related indicators in the U-Multirank pilot study, bibliometric data was derived from the October 2010 edition of the Web of Science bibliographical database. An upgraded 'bibliometric version' of the database is housed and operated by the CWTS (being one of the CHERPA Network partners) under a full license from Thomson Reuters. This dedicated version includes the 'standardized institutional names' of higher education and research institutes that have been checked ('cleaned') and harmonized in order to ensure that as many as possible of the Web of Science-indexed publications are assigned to the correct institution. This data processing of address information is done at the aggregate level of the entire 'main' organization (not for sub-units such as departments or faculties). All the selected institutions in the U-Multirank pilot study produced at least one Web of Science-indexed research publication during the years 1980-2010.

The Web of Science, being both an international and multidisciplinary database, has its pros and cons. The bulk of the research publications are issued in peer-reviewed international scientific and technical journals, which mainly refer to discovery-oriented 'basic' research of the kind that is conducted at universities and research institutes. There are relatively few conference proceedings in the Web of Science, and no books or

[23] Yet another database is *Google Scholar*. This is a service based on the automatic recording by Google's search engine of citations to any author's publications (of whatever type) included in other publications appearing on the worldwide web.

[24] See: http://thomsonreuters.com/products_services/science/science_products/a-z/web_of_science/

monographs whatsoever, hence, publications referring to 'applied research' or 'strategic research' are underrepresented. It has a relatively poor coverage of non-English language publications. The coverage of publication output is quite good in the medical sciences, life sciences and natural sciences, but relatively poor in many of the applied sciences and social sciences and particularly within the humanities. The alternative source of bibliographical information, Elsevier's Scopus database, is likely to provide an extended coverage of the global research literature in those underrepresented fields of science.

For the following six indicators selected for inclusion in the U-Multirank pilot test (see chapter 6) one can derive data from the CWTS/Thomson Reuters Web of Science database:

1. total publication output
2. university-industry joint publications
3. international joint publications
4. field-normalized citation rate
5. share of the world's most highly cited publications
6. regional joint publications

We note that this set includes four new performance indicators (#2, #3, #5, #6) that were specially constructed for U-Multirank and that have never been used before in any international classification or ranking.

4.2.3 Patent databases

As part of the indicators in the Knowledge Transfer dimension, U-Multirank selected the number of *patent applications* for which a particular higher education and research institution acts as an applicant and (as part of that) the number of *co-patents* applied for by the institution together with a private organization.

Data for the co-patenting and patents indicators may be derived from patent databases. For U-Multirank, patent data were retrieved from the European Patent Office (EPO). Its Worldwide Patent Statistical Database (version October 2009)[25], also known as PATSTAT, is designed and published on behalf of the OECD Taskforce on Patent Statistics. Other members of this taskforce include the World Intellectual Property Organisation (WIPO), the Japanese Patent Office (JPO), the United States Patent and Trademark Office (USPTO), the US National Science Foundation (NSF), and the European Commission represented by Eurostat and by DG Research.

25 This version is held by the K.U. Leuven (Catholic University Leuven) and was licensed to its ECOOM unit (Expertise Centrum O&O Monitoring).

The PATSTAT patent database is especially designed to assist in advanced statistical analysis of patent data. It contains patent data from over 80 countries; adding up to 70 million records (63 million patent applications and 7 million granted patents). The patent data are sourced from offices worldwide, including of course the most important and largest ones such as the EPO, the USPTO, the JPO and the WIPO. Updates of PATSTAT are produced every six months, around April and October.

PATSTAT is a relational database: 20 related tables contain information on relevant dates (e.g. of patent filing, patent publication, granting of patent), on patent applicants and inventors, technological classifications of patents, citations from patents to other documents, family links[26], etc. Updates of PATSTAT are produced twice a year.

4.2.4 Data availability according to EUMIDA

Like the U-Multirank project, the EUMIDA project (see http://www.eumida.org) collects data on individual higher education and research institutions. The EUMIDA project is meant to test whether a data collection effort can be undertaken by EUROSTAT in the foreseeable future. EUMIDA covers 29 countries (the 27 EU member states plus two additional countries: Switzerland and Norway) and investigates the data available from national databases in as far as these are held/maintained by national statistical institutes, ministries or other organizations. The EUMIDA project has demonstrated that a regular data collection by national statistical authorities is feasible across (almost) all EU-member states, albeit for a limited number of indicators – mostly of an input (instead of output-) type.

The EUMIDA and U-Multirank project teams agreed to share information on issues such as definitions of data elements and data sources, given that the two projects share a great deal of data (indicators). The overlap lies mainly in the area of data related to the inputs (or activities) of higher education and research institutions. A great deal of this input-related information is used in the construction of the indicators in U-Map. The EUMIDA data elements therefore are much more similar to the U-Map indicators, since U-Map aims to build *activity profiles* for individual institutions whereas U-Multirank constructs per*formance profiles*.

The findings of EUMIDA point to the fact that for the more research intensive higher education institutions, data for the dimensions of Education and Research are relatively well covered, although data on graduate careers and employability are sketchy. Some

[26] A patent family is a set of patents taken in various countries to protect a single invention (when a first application in a country – the priority – is then extended to other offices). In other words, a patent family is the same invention disclosed by a common inventor(s) and patented in more than one country (see: US Patent and Trademark Office: www.uspto.gov).

data on scientific publications is available for most countries. However, overall, performance-related data is less widely available compared to input-related data items. The role of national statistical institutes is quite limited here and the underlying methodology is not yet consistent enough to allow for international comparability of data.

Table 4-1 below shows the U-Multirank data elements that are covered in EUMIDA and whether information on these data elements may be found in national databases (statistical offices, ministries, rectors' associations, etc.). The table shows that EUMIDA primarily focuses on the Teaching & Learning and Research dimensions, with some additional aspects relating to the Knowledge Transfer dimension. Since EUMIDA never had the intention to cover all dimensions of an institution's activity (or its performance), it is only natural that dimensions such as International Orientation and Regional Engagement are less prominent in the project.

The table illustrates that information on only a few U-Multirank data elements is available from national databases and, moreover, what data exists is available only in a small minority of European countries. This implies, once again, that the majority of data elements will have to be collected directly from the institutions themselves.

Table 4-1: Data elements shared between EUMIDA and U-Multirank: their coverage in national databases

Dimension	EUMIDA and U-Multirank data element	European countries where data element is available in national databases
Teaching & Learning	relative rate of graduate unemployment	CZ, FI, NO, SK, ES
Research	expenditure on research	AT*, BE, CY, CZ*, DK, EE, FI, GR*, HU, IT, LV*, LT*, LU, MT*, NO, PL*, RO*, SI*, ES, SE, CH, UK
	research publication output	AT, BE-FL, CY, CZ, DK, FI, FR, DE, GR, HU, IE, IT, LV, LT, LU, NO, NL, PL, PT*, RO*, SK, SI, ES, SE*, CH, UK
Knowledge Transfer	number of spin-offs	BE-FL, FR*, GR, IT (p), PT (p), ES
	third party funding	CY, CZ, DE, IT, NL, NO, PL, PT, ES, CH
	patents	AT, BE-FL, CZ, EE*, FI, FR*, GR, HU, IE*, IT, LU, MT*, NO, NL (p), PL*, SI, ES, UK

International Orientation	(no overlap between U-Multirank and EUMIDA)	
Regional Engagement	(no overlap between U-Multirank and EUMIDA)	

- Source: Based on EUMIDA Deliverable D2 – *Review of Relevant Studies* (dated 20 February 2010 and submitted to the Commission on 1 March 2010).
- indicates: There are confidentiality issues (e.g. national statistical offices may not be prepared to make data public without consulting individual HEIs)
- (p) indicates: Data are only partially available (e.g. only for public HEIs, or only for (some) research universities)
- The list of EUMIDA countries with abbreviations: Austria (AT), Belgium (BE), [Belgium-Flanders community (BE-FL)], Bulgaria (BG), Cyprus (CY), Czech Republic (CZ), Denmark (DK), Estonia (EE), Finland (FI) France (FR), Germany (DE), Greece (GR), Hungary (HU), Ireland (IE), Italy (IT), Latvia (LV), Lithuania (LV), Luxembourg (LU), Malta (MT), Netherlands (NL), Norway (NO), Poland (PL), Portugal (PT), Romania (RO), Slovakia (SK), Slovenia (SI), Spain (ES), Sweden (SE), Switzerland (CH), United Kingdom (UK).

4.2.5 Expert view on data availability in non-European countries

The Expert Board of the U-Multirank project was consulted to assess for their six countries – all from outside Europe – the availability of data related to the U-Multirank indicators.[27] They gave their judgement on the question whether data was available in national databases and/or in the institutions themselves. Table 4-2 shows that the Teaching and Learning dimension scores best in terms of data availability. The dimensions Research and Knowledge Transfer have far less data available on the national level, but this is compensated by the data available at the institution level. The same holds true, to a lesser extent, for the dimension International Orientation, where little data is available in national databases. The Regional Engagement dimension is the most problematic in terms of data availability. Here, data will have to be collected from the individual institutions.

[27] Argentina, Australia, Canada, Saudi Arabia, South Africa and the US.

Table 4-2: Availability of U-Multirank data elements in countries' national databases according to experts in 6 countries (Argentina/AR, Australia/AU, Canada/CA, Saudi Arabia/SA, South Africa/ZA, United States/US)

Dimension	U-Multirank data element	Countries where data element is available in national databases	Countries where data element is available in institutional database
Teaching & Learning	expenditure on teaching	AR, US, ZA	AR, AU, SA, ZA
	time to degree	AR, CA, US, ZA	AR, AU, CA, SA, ZA
	graduation rate	AR, CA, US, ZA	AR, AU, SA, ZA
	relative rate of graduate unemployment	AU, CA, US	
Research	expenditure on research	AR, AU, ZA	AR, AU, SA, US, ZA
	number of post-doc positions		CA, US, ZA
	research publication output	AR, AU, US	AR, AU, SA, US, ZA
	international prizes won		AR, CA, ZA
Knowledge Transfer	size of TTO		AU, CA, SA, ZA
	incentives for knowledge exchange	AR	AR, AU, CA, SA
	CPD courses offered		AU, CA, SA, ZA
	university-industry joint research publications	AR	
	number of spin-offs	AU	CA, US
	third party funding	AU, US	CA, US, ZA
	license income		CA, US, ZA
	license agreements	AU	AR, CA, ZA
	co-patents		CA, ZA
	Patents	AR	AR, CA, US, ZA

International Orientation	educational programs in foreign language	ZA	AR, AU, CA, SA, ZA
	international academic staff	ZA, US	AR, AU, CA, SA, US, ZA
	joint degree programmes	AR	AR, AU, CA, US
	international doctorate graduation rate	US	AR, CA, SA, US
Regional Engagement	income from regional sources		AU, CA, SA, ZA
	student internships in local/regional enterprises		AU, SA, US, ZA
	graduates working in the region		US
	research contracts with regional business		AR, CA, ZA
	co-patents with regional firms	ZA	CA, ZA
	regional participation in continuing education		AR, CA, ZA

Source: Based on U-Multirank expert survey

If we look at the outcomes, it appears that for the Teaching & Learning indicators the situation is rather promising (graduation rate, time to degree). In the Research dimension, Expenditure on Research and Research Publication Output data are best represented in national databases. For the other dimensions, however, information is not really available in national databases. According to the experts consulted, more data can probably be found in institutional databases. However, if that is the case, there is always a risk that different institutions may use different definitions or definitions that differ from the ones used for the questionnaires applied in U-Multirank (see next section).

Even if there is information available in databases (national, institutional, or other), our experts stressed that it is not always easy to obtain that information (for instance in case of data relating to the dimension Regional Engagement). To obtain a better idea of data availability, we carried out a special pre-test (see section 4.3.3).

4.3 Data collection instruments

Due to the lack of adequate data sets, the U-Multirank project had to rely largely on self-reported data (both at the institutional and field-based levels), collected directly from the higher education and research institutions. The main instruments to collect data from the institutions were four online questionnaires: three for the institutions and one for students.

The four surveys are:

- U-Map questionnaire
- institutional questionnaire
- field-based questionnaire
- student survey.

In designing the questionnaires, emphasis was placed on the way in which questions were formulated. It is important that they can only be interpreted in one way and that the reasons for asking a question are evident. A mix of open and closed questions was used. Although a standard reference period is prescribed, we allowed for deviation in some cases. The questionnaires also contain the option for respondents to add comments to their answers. To facilitate completion of the questionnaires, answers can be saved temporarily for later access by the respondents. More detailed information on the questionnaires is provided in the sections below.

4.3.1 Self-reported institutional data

4.3.1.1 U-Map questionnaire

As explained, the U-Map questionnaire is an instrument for identifying similar subsets of higher education institutions within the U-Multirank sample. Data is collected in seven main categories:

- general information: name and contact; public/private character and age of institution;
- students: numbers; modes of study and age; international students; students from region;
- graduates: by level of program; subjects; orientation of degrees; graduates working in region;
- staff data: fte and headcount; international staff;
- income: total income; income by type of activity; by source of income;
- expenditure: total expenditure; by cost centre; use of full cost accounting;
- research & knowledge exchange: publications; patents; concerts and exhibitions; start-ups.

The academic year 2008/2009 was selected as the default reference year. Respondents from the institutions were advised to complete the U-Map questionnaire first before completing the other questionnaires.

4.3.1.2 Institutional questionnaire

By means of U-Multirank's institutional questionnaire[28], data is collected on the performance of the institution. Like the U-Map questionnaire, this questionnaire is structured along the lines of different data types to allow for a more rapid data collection by the institution's respondents. The questionnaire is therefore divided into the following categories:

- general information: name and contact; public/private character and age of institution; university hospital;
- students: enrolments;
- programme information: bachelor/master programmes offered; CPD courses;
- graduates: graduation rates; graduate employment;
- staff: fte and headcount; international staff; technology transfer office staff;
- income: total; income from teaching; income from research; income from other activities;
- expenditure: total expenditure; by cost centre; coverage;
- research & knowledge transfer: publications; patents; concerts and exhibitions; start-ups.

As the institutional questionnaire and the U-Map questionnaire partly share the same data elements, institutions were advised to first complete the U-Map questionnaire. Data elements from U-Map are automatically transferred to the U-Multirank questionnaire using a 'transfer tool'. The academic year 2008/2009 was selected as the default reference year.

4.3.1.3 Field-based questionnaire

The field-based questionnaire includes information on individual faculties/departments and their programmes in the pilot fields of business studies, mechanical engineering and electrical engineering. Like the institutional questionnaire, the field-based questionnaire is structured along the different types of data requested to reduce the administrative burden for respondents. Data was collected for the reference period 2009/2010 for data which are expected to be subject to annual fluctuations; data for three subsequent years was collected to calculate three-year averages.

[28] See Appendix 12 for the institutional questionnaire.

The following categories are distinguished:

- overview: name and address of unit responsible for organising the field; contact person;
- staff & PhD: academic staff; number of professors; international visiting/guest professors; professors offering lectures abroad; professors with work experience abroad; number PhDs; number post docs;
- funding: external research funds; license agreements/income; joint R&D projects with local enterprises;
- students: total number (female/international degree and exchange students); internships made; degree theses in cooperation with local enterprises;
- regional engagement: continuing education programmes/professional development programmes; summer schools/courses for secondary education students;
- description: accreditation of department; profile with regard to teaching & learning, profile with regard to research.

A second part of the questionnaire asks for details of the *individual study programmes* to be included in the ranking. In particular the following information was collected:

- basic information about the programme (e.g. degree, length); interdisciplinary characteristics; full time/part time;
- number of students enrolled in the programme; number of study places and level of tuition fees; periods of work experience integrated in programme; international orientation; joint study programme;
- credits earned for achievements abroad; number of exchange students from abroad; courses held in foreign language; special features;
- number of graduates; information about labor market entry.

4.3.2 Student Survey

For measuring student satisfaction (see section 3.3.1), the main instrument is an online student survey. In order to assure that students are not pressured by their institution/teachers to rate their own institution favorably, the institutions were asked to invite their students individually to participate in the survey either by mail or email – rather than having them complete the survey in the classroom. Access to the questionnaire was controlled by individual passwords. The student questionnaire uses a combination of open questions and pre-defined answers and asks for the students' basic demographic data and information on their programme. The main focus of the survey is on the assessment of the teaching and learning experience and on the facilities of the institution.

In order to control for possible manipulation by institutions, a number of control questions were included in the questionnaire. Students were asked for information on how they received the invitation and whether there were any attempts by teachers, deans or others to influence their ratings.

In relation to the student survey, the delimitation of the sample is important. As students were asked to rate their own institution and programme, students who had just started their degree programme were excluded from the sample. Hence students from the second year onwards in bachelor and master programmes and from the third year onwards in long (pre-Bologna) programmes were meant to be included. In order to have a sample size that allows for analysis, the survey aimed to include up to 500 students by institution and field.

4.3.3 Pre-testing the instruments

A first version of the three new data collection instruments (the institutional questionnaire, department questionnaire and student questionnaire) was tested between June and September 2010. The U-Map questionnaire had already been tested. The U-Multirank questionnaires were tested in terms of cultural/linguistic understanding, clarity of definitions of data elements and feasibility of data collection. Ten institutions were invited to complete and comment on the institutional and departmental questionnaire and to distribute 20 student questionnaires. The selection was based on the list of institutions that had expressed their interest in participating in the project. In selecting the institutions for the pre-test the U-Multirank team considered the geographical distribution and the type of institutions.

Since not all institutions responded fully to the pre-test, a 'light version' was sent to an additional 18 institutions. Instead of asking them to provide all the data on a relatively short notice, these institutions were contacted to offer their feedback on the clarity of the questions and on the availability of data.

According to the pre-test results, the general format and structure of the institutional questionnaire seemed to be clear and user-friendly. The pre-test showed, however, two types of problems for some indicators. Several indicators require a more precise specification, definition and/or examples. Respondents worried that for some indicators the definitions might not be sufficient for internationally comparable results. Secondly, several indicators presented difficulties to respondents because the required data was not centrally collected by the institution. Some of the frequently mentioned availability problems are presented separately for each dimension.

Teaching and learning. Questions about student numbers and study programmes seem to be unproblematic in most cases. Problems emerge however with some output-related

data elements such as graduate employment, where often data is not collected at the institutional level. Interdisciplinarity of programs proved to be another problematic indicator, where problems emerged due to the definition of the concept and the absence of the required data.

Research. Most data items in this dimension did not lead to problems. In fact, some of the key indicators are extracted from international bibliometric databases anyway and did not need data provision from the institutions. As expected, some difficulties emerged for 'art-related outputs'. Sharper definitions were called for here.

Knowledge Transfer and *Regional Engagement.* Compared to Teaching and Research, these two dimensions are less prevalent in existing national and institutional databases and therefore presented some data availability problems. This was the case for 'graduates working in the region' and 'student internships in regional enterprises'. Comprehensive information on start-up firms and professional development courses was not always available for institutions as a whole.

International Orientation. Information on international students and staff, as well as on programmes in a foreign language was largely available. As expected, the question of how to define an 'international student' came up occasionally.

In sum, the institutional questionnaire worked well in terms of its structure and usability. The respondents did not find the questionnaire excessive or burdensome. The pre-test did reveal a need for clearer definitions for some data elements. Pre-test results also indicated that some data elements, although highly relevant and valid, could not be feasibly collected because institutions did not have such data. With respect to this issue the project team, with the help of the Advisory Board, had a critical look at the problematic indicators and decided which items to drop and which to keep in the further stages in the project. In short, the feedback and recommendations from the pre-test were fed into a second, modified version of the institutional questionnaires that were used during the pilot phase.

The field-based questionnaire was pre-tested in five departments. Some other institutions sent a few general comments on particular issues and questions. Problems with regard to the availability of data were reported mainly on issues of academic staff (e.g. fte data, international staff), links to business (in education/internships and research) and the use of credits (ECTS). The definition of the categories of academic staff ('professors' – 'other academic staff') clearly depends on national legislation and definitions.

The length of the questionnaire was raised as an issue. Some institutions wished to have a shorter questionnaire, yet some mentioned missing items such as those relating to e.g. social issues and diversity.

The student survey was pre-tested on a sample of over 80 students. In general, their comments were very positive. The questions were felt to be clear and understandable and captured relevant issues of the students' teaching and learning experience/environment. Some students would have preferred more questions about the social climate at the institution and about the city or town in which it was situated; a number of reactions (also from pre-test institutions) indicated that the questionnaire should not be any lengthier, however.

Comments were received about the phrasing of some questions – in particular the need to take the national structures and situations into account. A major challenge deduced from these comments is how to compare across cultures students' assessment of their institutions. Based on approved instruments from other fields (e.g. surveys on health services) we have used 'anchoring vignettes' to test socio-cultural differences in assessing specific constellations of services/conditions in higher education with respect to teaching and learning. These anchoring vignettes are explained in section 6.3.2 and appendix 9.

The main conclusions from the pretest were:

- The project had to find a compromise between two conflicting goals: to cover all relevant issues on the five dimensions of U-Multirank and to limit the questionnaire in terms of length. A particular problem of the study was that we could not decide *a priori* which indicator would be valid, reliable and feasible. In order to come to a meaningful and comprehensive set of indicators at the conclusion of the U-Multirank pilot study we had to aim for a broad data collection to cover a broad range of indicators. The final list of indicators that came out of the wider pilot test (presented in chapter 6) was assumed to be less extensive than the list analyzed in the pre-test phase.
- One will have to deal with the issue of institutions providing 'estimated' values instead of data from existing data sets. In the pre-test, institutions were allowed to provide estimates as long as they clearly indicated such cases; this enabled us to get an impression about the precision of data.
- For the student questionnaire the conclusion was that there is no need for changes in the design. Comments received showed that the questionnaire is seen as a useful instrument.

4.3.4 Supporting instruments

In order to assure that a comparable data set was obtained, a number of supporting instruments were prepared for the four U-Multirank surveys. These instruments ensure that respondents will have a common understanding of definitions and concepts. This is particularly important as institutions from diverse national settings are an important source for data collection. The pre-test indicated differences between countries in terms of understanding some of the items. The following supporting instruments were provided to offer more clarity to the respondents during the process of data collection:

- A *glossary* of indicators for the four surveys was published on the U-Multirank website. Throughout the data collection process the glossary was updated regularly.

- A '*frequently asked questions*' (FAQ) section next to a 'Helpdesk' function was launched on the website. This allowed questions to be asked concerning the questionnaires and for contact with the U-Multirank team on other matters.

- P*rotocols* describing data collection and handling were developed to explain to the institutions in detail how the different steps were laid out from the start to the finish and the finalising of the data collection.

- A *technical specifications protocol* for U-Multirank was developed, introducing additional functions in the questionnaire to ensure that a smooth data collection could take place: the option to download the questionnaire in Pdf format, the option to transfer data from the U-Map to the U-Multirank institutional questionnaire, and the option to have multiple users access the questionnaire at the same time.

- We updated the U-Multirank website regularly and provided information about the steps/time schedules for data collection.

- All institutions had clear communication partners from the U-Multirank team.

4.4 A concluding perspective

This chapter, providing a quick survey of existing databases, underlines that there are very few international databases/sources where data can be found for our type of rankings. The only sources that are available are international databases holding bibliometric and patent data. This implies that, in particular for a ranking that aims to sketch a multidimensional picture of an institution at the institutional and disciplinary field levels, one will have to rely to a large extent on data collected by means of questionnaires sent to representatives of institutions, their students and – possibly – their

graduates. One could even go beyond these stakeholder groups and include employers and other clients of higher education and research institutions, but that would make the task even bigger. The way the data are collected then becomes a critical issue, where compromises have to be built between comprehensiveness, feasibility, respondents' efforts, etc. Different questionnaires will have to be sent to the different data providers: institutions, representatives of departments in the institution and students. Sampling techniques (selecting/identifying institutions, departments/programmes, their representatives and their students) are crucial, as is the intelligent use of technology (internet, visualisation techniques, supporting tools). The language of the questionnaire is another crucial element for ensuring a good response to the questionnaire.

In addition, challenges in terms of comparability occur. As rankings order their objects in terms of their scores on quantitative indicators they require uniform definitions of the underlying data elements. The U-Multirank questionnaires therefore were accompanied by a glossary of definitions and an FAQ facility to improve the reliability of the answers. However, as a result of differences in national higher education systems, different accounting systems, as well as different national customs and definitions of indicators, there are limits to the comparability of data. Therefore, respondents will always have to have the opportunity to provide footnotes and comments to the data they submit through the questionnaires. In a few cases, one may have to allow respondents to provide estimates for some of the answers if data is otherwise unavailable or too costly to collect. Checking the answers can be done based on internal consistency checks, comparing data to that of other institutions, or making use of data from other sources, but this clearly also has its limits.

What this chapter has made clear is that the questionnaires and surveys need to be tested first on a small scale before embarking on a bigger survey. Taking into account the experiences from other similar ranking/data collection projects, and making use of the advice of external experts and national correspondents in the testing and further execution of the survey is yet another part of the provision that needs to be part of the data collection strategy.

5 Testing U-Multirank: pilot sample and data collection

5.1 Introduction

Now that we have presented the design and construction process for U-Multirank, we will describe the feasibility testing of this multidimensional ranking tool. This test took place in a pilot study specifically undertaken to analyse the actual feasibility of U-Multirank on a global scale. In this chapter we will describe the processes of recruiting the sample of pilot institutions and data collection in the pilot study – the collection of both self-reported institutional data and data from international databases.

5.2 The global sample

A major task of the feasibility study was the selection of institutions to be included in the pilot study. The selection of the 150 pilot institutions (as specified in the project outline) needed to be informed by two major criteria: including a group of institutions that reflects as much institutional diversity as possible; and making sure that the sample was regionally and nationally balanced. In addition we needed to ensure sufficient overlap between the institutional ranking and the field-based rankings in business studies and two fields of engineering.

As has been indicated in chapter 2 of this report, one of the basic ideas of U-Multirank is the link to U-Map. U-Map is an effective tool to identify institutional activity profiles of institutions similar enough to compare them in rankings. Yet at this stage of its development U-Map includes only a limited number of provisional institutional profiles which makes it insufficiently applicable for the selection of the sample of pilot institutions for the U-Multirank feasibility test. Since U-Map cannot yet offer sets of comparable institutional profiles we needed to find another way to create a sample with a sufficient level of diversity of institutional profiles. We do not (and cannot) claim that we have designed a sample that is representative of the full diversity of higher education in the world (particularly as there is no adequate description of this diversity) but we have succeeded in including a wide variety of institutional types in our sample.

Potential pilot institutions to be invited for the sample were identified in a number of ways:

- The existing set of higher education institutions in the U-Map database was included. This offered a clear indication of a broad variety of institutional profiles.

- Some universities applied through the U-Multirank website to participate in the feasibility study. Their broad profiles were checked as far as is possible against the U-Map dimensions in order to be able to describe their profiles.
- In most countries 'national correspondents' (a network created by the research team) were asked to suggest institutions that would reflect the diversity of higher education institutions in their country. Clearly this is easier to do in large countries where we planned to include six or more institutions (see Table 5-1) than in small countries where only one or two institutions could be included. For the latter countries we looked at institutional diversity across the group of small countries.
- Some international networks of institutions expressed an interest to be involved in the project and suggested institutions with specific profiles to participate in the pilot study.
- Our field-based partner organizations (FEANI, EFMD) were consulted with regard to the field based rankings and suggested institutions that offer programmes in the fields addressed by the pilot study (business studies and two fields of engineering).

Looking at the final sample, we are confident that the group of pilot institutions has sufficient institutional diversity. The U-map profiles of the institutions reflect variation regarding the five dimensions. In addition the sample covers a few specialized institutions. To illustrate this, the sample includes: an Institute for Water and Environment, an agricultural university, a School of Petroleum and Minerals, a military academy, several music academies and art schools, universities of applied sciences and a number of technical universities.

The 159 institutions that agreed to take part in the U-Multirank pilot are spread over 57 countries. The distribution between European and non-European countries is as follows: 94 institutions are from countries of the European Union; 15 are from non-European Union but European countries and 50 institutions are from outside Europe. This distribution reflects a 2/3 mix between European and non-European countries.

Two countries turned out to be particularly problematic: the US and China. Our national correspondents explained that Chinese universities are reluctant to participate in rankings when they cannot predict the outcomes of participation and fear being placed in an unfavourable position. In addition there appear to be formal reasons why the Chinese government is hesitant to stimulate participation. In the US the U-Multirank project is perceived as strongly European-focused, which kept some institutions from participating. For both countries we tried to address and resolve these concerns, and contacted our networks again, published articles in relevant newsletters and received help from the European Commission. Yet it was impossible to reach the target number in

these two countries. On the other hand there was an interest from regions/countries that were not initially intended to be included, i.e. Africa, Latin America and the Near East.

The problems with some countries are an important aspect regarding the feasibility of a global implementation of U-Multirank. All in all the intention to attain a sufficient international scope in the U-Multirank pilot study by means of a global sample can be seen as successful. Finally 115 institutions submitted data as part of the pilot study.

Table 5-1: Regional distribution of participating institutions

Region and Country	Initial proposal for number of institutions	Institutions in the final pilot selection July 2010	Institutions that confirmed participation February 2011	Institutions which delivered U-Multirank institutional data April 2011	Institutions which delivered U-Multirank institutional data and U-Map data April 2011
I. EU 27 (population in millions)					
Austria (8m)	2	2	5	**5**	**4**
Belgium (10m)	3	3	5	**3**	**3**
Bulgaria (8 m)	2	3	3	**3**	**3**
Cyprus (1m)	1	1	1	**1**	**0**
Czech Republic (10m)	3	4	4	**4**	**4**
Denmark (5m)	2	5	4	**4**	**3**
Estonia (1m)	1	2	1	**1**	**1**
Finland (5m)	2	3	2	**2**	**2**
France (64m)	6	9	6	**3**	**3**
Germany (84m)	6	9	8	**5**	**5**
Greece (11m)	3	4	2	**1**	**1**
Hungary (10m)	3	4	3	**3**	**3**
Ireland (4m)	1	1	6	**5**	**5**
Italy (60m)	6	8	6	**5**	**5**
Latvia (2m)	1	1	1	**1**	**1**
Lithuania (3m)	1	2	1	**1**	**1**
Luxembourg (0.5m)	1	1	1	**0**	**0**
Malta (0.4m)	1	1	0	**0**	**0**

Region and Country	Initial proposal for number of institutions	Institutions in the final pilot selection	Institutions that confirmed participation	Institutions which delivered U-Multirank institutional data	Institutions which delivered U-Multirank institutional data and U-Map data
Netherlands (16m)	3	7	3	**3**	**3**
Poland (38m)	6	12	7	**7**	**6**
Portugal (10m)	3	3	4	**3**	**3**
Romania (21m)	3	5	5	**4**	**4**
Slovakia (5m)	2	1	1	**0**	**0**
Slovenia (2m)	1	2	1	**1**	**1**
Spain (46m)	6	7	7	**5**	**4**
Sweden (9m)	2	3	3	**3**	**3**
United Kingdom (62m)	6	8	4	**2**	**2**
Total EU	*77*	*102*	*94*	75	70
II. Europe – Non EU					
Russia	5	4	2	**2**	**2**
Switzerland		6	5	**3**	**3**
Turkey		6	2	**2**	**2**
Norway		4	4	**3**	**3**
Liechtenstein		1	0	**0**	**0**
Iceland		1	1	**1**	**1**
Croatia	-	-	1	**1**	**1**
Total non-EU	*5*	*22*	*15*	***12***	***12***
III. Outside Europe					
US	19	24	4	**1**	**1**
Canada	6	6	3	**1**	**1**
Japan	5	9	2	**2**	**2**
China	10	11	2	**1**	**1**
India	5	7	4	**2**	**2**
Australia	3	8	7	**6**	**6**

Region and Country	Initial proposal for number of institutions	Institutions in the final pilot selection	Institutions that confirmed participation	Institutions which delivered U-Multirank institutional data	Institutions which delivered U-Multirank institutional data and U-Map data
Other Asia	5	2			
• **The Philippines**			1	**1**	**1**
• **Taiwan**			1	**1**	**0**
• **Vietnam**			2	**1**	**1**
• **Malaysia**			1	**0**	**0**
• **Indonesia**			1	**0**	**0**
Latin America	5	3			
• **Mexico**			2	**2**	**2**
• **Colombia**			1	**1**	**1**
• **Chile**			1	**1**	**1**
Africa	5				
• **South Africa**		5	3	**0**	**0**
• **Other Africa**		3	3	**1**	**1**
Israel	5	2	2	**1**	**1**
Saudi Arabia		4	3	**2**	**2**
Other Middle East		1			
• **Algeria**			2	**2**	**2**
• **Lebanon**			1	**1**	**1**
• **Tunisia**			1	**1**	**1**
• **Egypt**			1	**0**	**0**
• **Morocco**			2	**0**	**0**
Total non-Europe	*68*	*75*	*50*	***28***	***27***
Total	***150***	***216***	***159***	***115***	***109***

During the pilot study there was some criticism that top research institutions were underrepresented in our sample. For this reason we performed an additional check on the representativeness of our sample in terms of the inclusion of internationally-oriented top research institutions. We analysed how the institutions of our sample perform in existing international rankings focusing on research excellence. The analyses showed that a significant number of institutions in our sample are listed: 19 institutions are in the top 200 of the Times Higher Education ranking, 47 in the top 500 of the ARWU ranking and 47 in the top 500 of the QS ranking. Since the exact number of higher education institutions in the world is not known we use a rough estimate of 15,000 institutions worldwide. In that case the top 500 comprises only 3% of all higher education institutions. In our sample 29% of the participating institutions are in the top 500, which indicates an overrepresentation rather than an underrepresentation of research intensive institutions in our sample.

With respect to the sample at the level of the three fields of study the situation was as follows. Of the 272 departments that agreed to participate in the field-based pilot study 165 (61%) (partially) completed the departmental questionnaire. Participation across the three fields was well-balanced: 57 departments in business studies participated, 50 in electrical engineering and 58 in mechanical engineering. Many institutions participated in more than one field and 14 did so in all three fields.

As has been explained, the field pilot study included a student satisfaction survey. Participating institutions were asked to send invitations to their bachelor and master's students to take part in a survey. 106 departments agreed to do so. Some institutions decided to submit the information requested in the departmental questionnaire but not to participate in the student survey as they did not want it to compete with their own surveys or effect participation in national surveys. In other institutions, students were on holiday or taking examinations during the pilot study survey window. In some cases the response rate was very low and the institutions concerned were excluded from this dimension of the analysis.

In total 6,770 students provided data via the online questionnaire. After data cleaning we were able to include 5,901 student responses in the analysis: 45% in business studies; 23% in mechanical engineering; and 32% in electrical engineering.

5.3 Data collection

The data collection for the pilot study took place via two different processes: the collection of self-reported data from the institutions involved in the study (including the student survey) and the collection of data on these same institutions from existing international databases on publications/citations and patents. In the following sections we discuss these data collection processes.

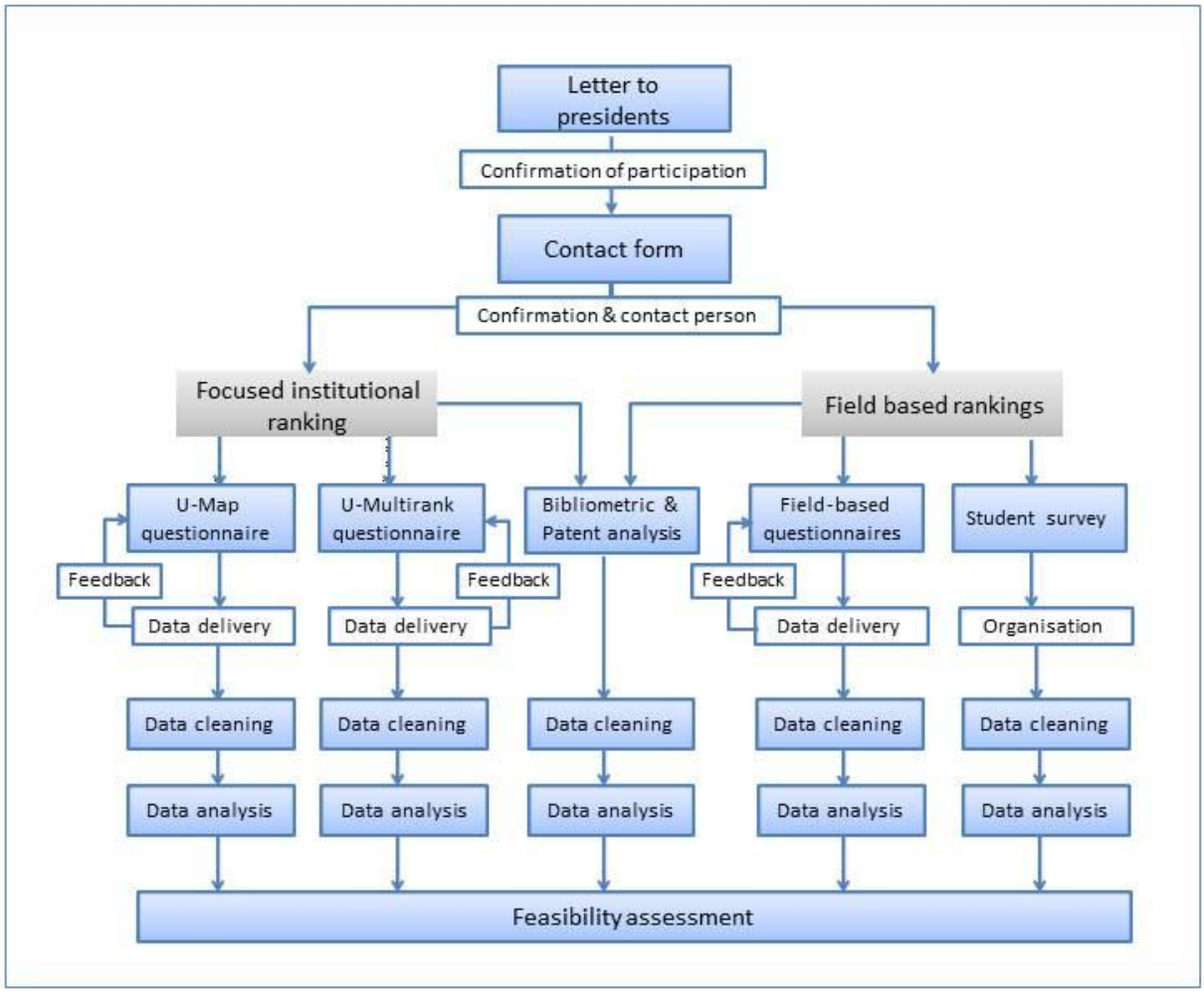
5.3.1 Institutional self-reported data

5.3.1.1 The process

The process of data collection from the organizations was organised in a sequence of steps (see Figure 5-1). First we asked the institutions, after official confirmation of participation, to fill in a contact form. This contact form identified a person at the institution as the contact for the project. This contact person received the access codes for the questionnaires. When an institution did not fill in the contact form we sent a reminder after two weeks. If we did not receive an answer we called the institution. This follow-up call was in nearly all cases regarded as valuable and most of the institutions submitted their contact form in response. The data collection entailed the following instruments:

- The U-Map questionnaire to identify institutional profiles
- Institutional ranking:
    - U-Multirank institutional questionnaire
- Field-based ranking:
    - U-Multirank field-based questionnaires
    - U-Multirank Student survey

Figure 5-1: U-Multirank data collection process



The institutions were given seven weeks to collect the data, with deadlines set according to the dates the institution confirmed their participation. The 'grouping' criterion for this was the successful submission of the contact form.

The next step to ensure a high response rate was to review whether the institutions did in fact access their questionnaire accounts. If the account had not been accessed four weeks after sending out the access information, we emailed a reminder and asked whether there had been any problem with the account. We advised the institutions to start working with the questionnaires in a certain order beginning with the U-Map and then the U-Multirank questionnaires, since a tool had been developed to facilitate the transfer of overlapping information from the U-Map questionnaire to the U-Multirank institutional questionnaire. The field-based questionnaires could be completed in parallel

to the institutional questionnaires. After the deadlines for data submission had passed, we checked on the questionnaires submitted by the institutions. If only one of the two institutional questionnaires had been submitted, reminders were sent out. These different steps allowed us to actively follow the data collection process and to assist institutions as needed.

An important element in terms of quality assurance of the data was a feedback loop built into the process. After the institutions had submitted their questionnaires their data was checked and we provided comments and questions. This provided the institutions with an opportunity for a second submission in which they could provide answers to the questions, check their data, correct inconsistencies and add missing information.

Organising a survey among students on a global scale was one of the major challenges in U-Multirank. There are some international student surveys (such as 'Eurostudent') but these usually focus on general aspects of student life and their socio-economic situation. To the best of our knowledge there is no global survey asking students to assess aspects of their own institutions and programmes. So we had no way of knowing whether students from different countries and cultures would assess their institutions in comparable ways. In Chapter 8 (8.2) we will discuss the flexibility of our approach to a global scale student survey.

The data collection through the student survey was organized by the participating institutions. They were asked to send invitation letters to their students, either by regular mail or by email. We prepared a standard letter to students explaining the purpose of the survey/project and detailing the URL and personal password they needed to access the online questionnaire. Institutions were able to download a package including the letter and a list of passwords (for email invitation) and a form letter (for printed mail invitations). If the letters were sent by post, institutions covered the costs of postage. No institution indicated that it did not participate in the student survey because of the cost of inviting the students.

In some countries (e.g. Australia) the students were taking examinations or were on vacation at the time the survey started. As a consequence some institutions decided not to participate in the survey; others decided to postpone the survey. As indicated earlier a total of 6,770 students participated in the survey, of this total 5,901 could be included in the analysis.

5.3.1.2 *Follow-up survey*

After the completion of the data collection process we asked those institutions that submitted data to share their experience of the process and to provide comments or suggestions for further improvement of the procedures and instruments. In this section we highlight the main outcomes.

One particular issue was the burden of data delivery in the various surveys. As can be seen in Table 5-2 this burden differed substantially between the pilot institutions. The average time spent per questionnaire was around five to six days.

Table 5-2: Self-reported time needed to deliver data (fte staff days)

Data collection tool	N	Minimum	Maximum	Mean
Institutional questionnaire	26	1.0	30	6.9
Field questionnaire Business	14	0.2	20	4.7
Field questionnaire Electrical Engineering	11	1.0	15	5.5
Field questionnaire Mechanical Engineering	14	1.0	20	6.0
Organization of student survey	18	0.2	21	4.4

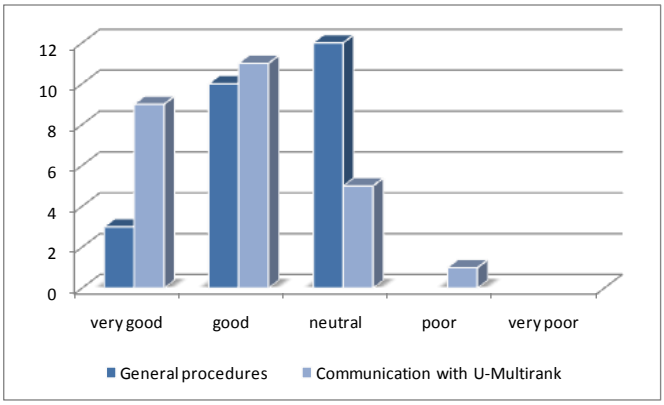
The analysis also showed that European institutions spent significantly less time on delivering the data than the institutions from outside Europe.

Table 5-3: Self-reported time needed to deliver data (fte staff days): European vs. non-European institutions

Data collection tool	Europe		Non-Europe	
	Mean	N	Mean	N
Institutional questionnaire	6.2	15	8.3	10
Field questionnaire Business Studies	2.5	10	7.3	7
Field questionnaire Electrical Engineering	3.5	8	7.0	5
Field questionnaire Mechanical Engineering	4.6	7	7.0	4
Organization of student survey	4.1	7	7.9	7

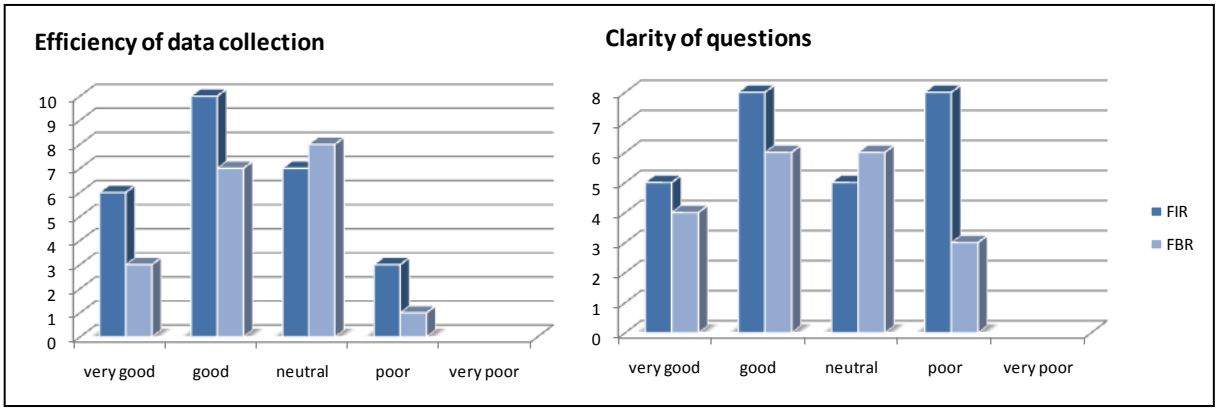
Figure 5-2 shows that the data collection process and procedures were judged positively by pilot institutions although some institutions were not completely satisfied.

Figure 5-2: Follow up survey: assessment of data procedures and communication



Other questions in the follow-up survey referred to the efficiency of data collection and the clarity of the questionnaires. In general the efficiency of data collection was reported to be good by the pilot institutions; critical comments indicated some confusion about the relationship between the U-Map and U-Multirank institutional questionnaires.

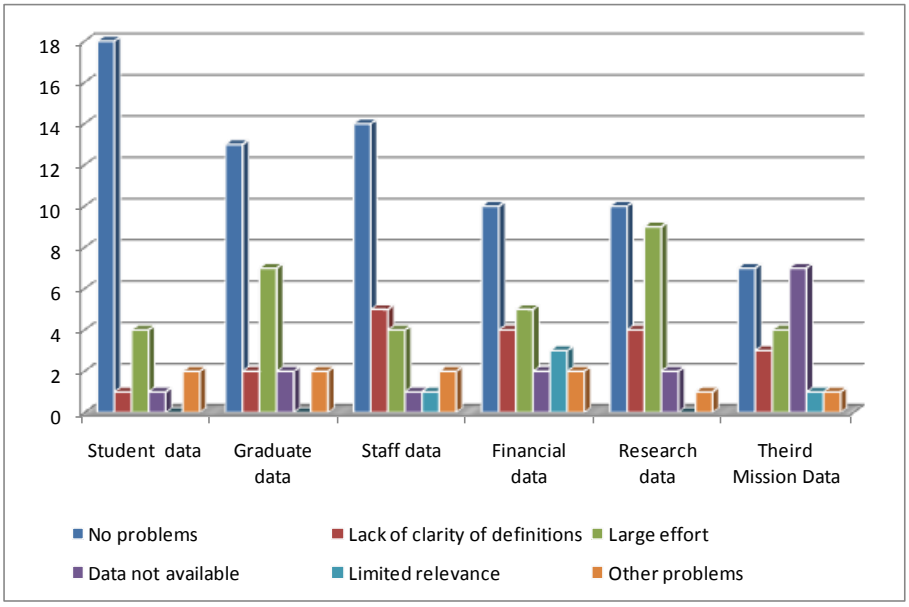
Figure 5-3: Follow up survey: assessment of data collection process



Some institutions were critical about the clarity of questions. Comments show that this criticism refers mainly to issues concerning staff data (e.g. the concept of full-time equivalents) and to aspects of research and knowledge transfer (e.g. international networks, international prizes, cultural awards and prizes).

In the follow-up survey we also asked about major problems in delivering the data. Most pilot institutions reported no major problems with regard to student, graduate and staff data. If they had problems these were mostly with research and third mission data (knowledge transfer, regional engagement) (See Figure 5-4).

Figure 5-4: Follow up survey: Availability of data



5.3.1.3 *Data cleaning*

As was indicated earlier, due to the lack of relevant and useful data sets we had to rely largely on self-reported data (both at the institutional and the field-based level). This inevitably raises the question of the control and verification of data. Based on the experiences from U-Map and from the CHE ranking we applied a number of mechanisms and procedures to verify data. Verification refers to the identification and correction of errors due to:

- Misunderstandings of definitions, concepts, etc.
- Simple data errors
- Potential manipulation of data

In order to reduce the number of errors due to misunderstanding of definitions and concepts we prepared a glossary of indicators for the four surveys. In addition to that a 'frequently asked questions' (FAQ) section and Helpdesk function were launched on the website. Furthermore, we shared the U-map protocol and the U-Multirank technical specification email (see appendices 10 and 11) with the institutions to ensure that a smooth data collection could take place. If despite these tools questions of definition still occurred, all universities had clear communication partners in the U-Multirank team.

The main part of the verification process consisted of the data cleaning procedures after receiving the data. A general and central feature of these procedures was the direct communication with the institutions. If inconsistencies and questions could not be solved, the particular data were not included in the pilot data analysis. The main data

cleaning procedures carried out on the data provided by the institutions are described below.

The institutional questionnaires

For the institutional questionnaires we performed the following checks:

- A check on the outliers in the data elements: the raw data (the answers provided by the institutions) were first analysed regarding outliers. If a score was extremely high or low (compared to the scores of the other institutions on that data element), the data element was flagged for further analysis.
- A check on the outliers in indicator scores: the scores on the indicators were calculated using the raw data and the formulas. If a score was extremely high or low (compared to the scores of the other institutions on that indicator), the data element was flagged for further analysis.
- A check for missing values: the data elements where data were missing or not available were flagged. Comments regarding reasons for missing data were studied and the missing values were compared to data from other institutions from the same country.

These three checks were first performed for the entire data set. In addition, more detailed checks were performed within a country or region. The focus of these more detailed checks was on:

- Reference years: a basic check on the consistency of the reference years.
- Comments: the comments were used as a source of information for missing values and for potential threats to the validity due to deviant interpretations.

If an outlier occurred, the website of the institution was checked to see whether we could find information regarding the relevant data element. The same procedure was followed when information was missing. If the website did not provide the information, other publicly available data sources were identified and studied to find out whether the outlier was due to inadequate interpretation and data provision regarding the question/data element or to a particular characteristic of the institution.

The departmental questionnaires

For the departmental questionnaires the following checks took place:

- Feedback cycles during the data collection process. After the first deadline we reviewed the data delivered thus far and inserted questions into the questionnaire which was sent again to the institutions.
- Analyses of outliers: for each indicator outliers were identified and analysed in more detail. Points of reference were the total average of scores of a given

number/indicator, field-based averages, and national averages (as far as the sample included several institutions from that country).
- Analyses of differences within a country: as far as the sample allowed, an analysis took place to identify country-specific outliers or inconsistencies.
- Analyses of trends over time: most indicators refer to three-year averages. The data provided were studied over time and specific changes in trends were analysed.

The student survey

For the student survey, after data checks we omitted the following elements from the gross student sample:

- Missing data on the students' institution
- Missing data on their field of study (business studies, mechanical engineering, electrical engineering)
- Students enrolled in programs other than bachelor/short national first degree programs and master/long national first degree programs
- Students who had spent little time on the questionnaire and had not responded adequately. Students had to answer at least parts of the questions that are used to calculate indicators and give the necessary information about their institution, field of study and programme they are enrolled in. In addition we used the time to complete the questionnaire (which is tracked in the online survey system) as an indicator.
- Students who reported themselves as formally enrolled but not studying actively
- Students reporting that they had just moved to their current institution
- Students who obviously did not answer the questionnaire seriously
- In addition we performed a recoding exercise for those students who reported their field of study as 'other'. Based on their explanation and on the name of the programme they reported, the field was recoded manually in all cases where a clear attribution was possible. In this process we took into consideration the attribution of programmes to fields as reported by the institutions in the department questionnaire.

As a result of these checks the data of about 800 student questionnaires have been omitted from the sample.

5.3.2 International databases

The data collection regarding the bibliometric and patent indicators took place by studying the relevant international databases and extracting from these databases the information to be applied to the institutions and fields in the sample.

5.3.2.1 Bibliometric data

As indicated in chapter 4, we analysed the October 2010 edition of the Web of Science database (WoS) to compile the bibliometric data of the institutions involved in the sample. A crucial aspect of this analysis was the identification of the sets of publications produced by one and the same institution, which is then labelled by a single, 'standardised' name tag.

The institutions were delimitated according to the set of WoS-indexed publications that contain an author affiliate address explicitly referring to that institution. The address information may comprise full names, name variants, acronyms or misspellings. This information was – as yet - gathered in a 'top-down' manner, i.e. without an external 'bottom-up' verification of the addresses or publications. Such a verification process would need an interaction with one or more representatives of each institution. As a result, 100% completeness for the selected set of publications cannot be guaranteed.

The identified institutions may comprise multiple affiliations (branches) – including hospitals, clinics or other medical centers – located elsewhere within the same city, region or country. For the institutions participating in the sample, statistics were produced that are sufficiently represented in the WoS database, either in the entire WoS or in the pre-selected WoS fields of science.

Six indicators were selected for usage in either the institutional ranking and/or the field-based ranking. Note that this set includes four new performance indicators that have never been used before in any international ranking of higher education institutions. The following four indicators have especially been designed for U-Multirank:

- International joint research publications;
- University-industry joint research publications;
- Regional joint research publications;
- Highly cited research publications.

Further information on each of the six bibliometric indicators used in the pilot study is presented below.

1) Total publication output

Frequency count of research publications with at least one author address referring to the selected main organization. This is primarily an indicator of research output, reflecting research capabilities and capacity. Since these publications are issued in peer-reviewed journals, they also signify a certain degree of research quality.

2) International joint research publications

Frequency count of publications with at least one author address referring to the selected main organization and one or more other addresses referring to another country. This is an indicator of research collaboration with partners located in other countries.

3) University-industry joint research publications

Frequency count of publications with at least one author address referring to the selected main organization and one or more other addresses referring to a private sector organization. This is an indicator of research collaboration with partners in the private sector, either domestically or located in other countries. The delimitation of private sector organization was done in accordance to a CWTS classification system of institutional addresses into major institutional sectors (for more details, see Tijssen et al. 2009). Statistical information on 500 universities worldwide is freely available at the CWTS website: www.socialsciences.leiden.edu/cwts/products-services/scoreboard.html

4) Regional joint research publications

Frequency count of publications with at least one author address referring to the selected main organization and one or more other addresses referring to a private sector organization. This is an indicator of research collaboration with partners within the sub-national region. The delimitation of regions was done according to EUROSTAT's NUTS system within Europe. In this pilot study the regions are NUTS2 regions, basically equivalent to provinces within most countries (the smallest European countries have no NUTS2 regions. This analysis is, by necessity, restricted to European main organizations.
In a possible next stage of U-Multirank we expect to apply a different, and more flexible, way of delineating regions which will enable us to broaden the scope beyond Europe. We will use the physical distance between co-authoring partners as computed from the distance between corresponding cities mentioned in the author addresses. The methodology was developed by CWTS in 2011 (Tijssen et al., 2011; Waltman et al., 2011). This will allow us delineate a series of regions, for example 'close vicinity regions' (where distances are less than say 10 kilometers), 'city regions' (11-50 kilometers) to 'longer distance regions' (50-100 kilometers).

5) Field-normalized citation rate

This is an indicator of citation-based international scientific impact. More specifically, a field-normalized citation impact score, where the fields are equivalent to the Thomson Reuters Journal Categories. We compare 'actual' citation counts to 'expected' counts based on the average impact score of all journals assigned to a field. A score larger than one represents a citation impact above world average within than field of science, whereas scores below one represent below average impact. Mean Normalized Citation Score scores (MNCS) between 0.8 and 1.2 are considered 'world average'; 1.2 to 1.5 is 'good' at the international level, and scores above 1.5 are associated with an 'excellent'

research performance. High scores on this indicator are associated with international-level 'research quality'.

6) Highly cited research publications

This is an additional indicator of citation-based international scientific impact that focuses on the top-end of the citation impact distribution within fields of science. The actual number of publications of a main organization within the world's top 10% most highly cited publication per field, is compared to the expected number of publications (i.e. 10% of organization's publication output in that same field).
We compare 'actual' citation counts to 'expected' counts per field: a score larger than one represents a 'surplus' of highly cited publications; a score below one a 'deficit'. A large surplus is associated with international-level 'research excellence'.

The bibliometric data in the pilot version of U-Multirank database refer to one measurement per indicator. In the case of the indicators #1-#4 (see section 4.2.2) the most recently available publication year was selected for producing statistical data: 2010. The statistics are in form of frequency data or as frequency categories (frequency range). Also, in the case of indicators #2, #3, and #4 the data were expressed as the share of co-publications within total publication input. The citation impact data require a citation window stretching back into the recent past in order to collect a sufficiently large number of citations. The window comprises the time-span 2005-2009 or 2006-2010.

The publication count data are all based on a 'whole counting' method where a publication is attributed in full to each main organization listed in the author addresses. The research publication counts refer to the following 'research-based' document types within the WoS: articles, notes, reviews, conference proceedings, papers, letters. The annual statistics refer to publication years (rather than database years).

The computation routine for the field-normalized citation rate indicator involved collecting citations to each publication according to a variable citation window, where each publication is tracked with the constraints of the pre-set time period. For instance, within the time period 2005-2009 all publications from 2005 are tracked for five years up to and including 2009; those published in 2006 were tracked for four years, etc. The most recent publication year was not included to prevent the occurrence of statistical biases in the field-normalized citation rate due to low citation counts and extremely low expected counts.

In the case of 'highly cited research publications' indicator, the citation impact distributions were calculated by applying a fixed citation window, for two 'research-based' document types: articles, reviews. These data refer to database years.

The research publications in the three fields of our pilot study (business studies, mechanical engineering and electrical engineering) were delimitated according to the WoS-indexed journal in which they are published, which are in turn classified by Thomson Reuters experts into one or more Journal Categories. The Journal Categories, sometimes referred to as Subject Categories, are treated as (sub)fields of science. Obviously, these fields should be seen as crude general representations of the corresponding knowledge domains. As such they may not (fully) align with the perceptions or institutional delineations of such a field within a main organization. These three fields comprise the following Journal Categories:

- Business: 'Business', 'Management', 'Business, Finance';
- Mechanical Engineering: 'Engineering, Mechanical', 'Engineering, Industrial';
- Electrical Engineering: 'Engineering, Electrical and Electronic'.

The fields in the institutions were delimitated according to the set of WoS-indexed publications that contain an author affiliate address explicitly referring to that institution. The address information may include full names, name variants, acronyms or misspellings. This information was – again - gathered in a 'top-down' manner, i.e. without an external 'bottom-up' verification of the addresses or publications.

With respect to the bibliometric analysis of our sample a final specific remark should be made. Although all the HEIs that participated in the U-Multirank pilot study produced at least one WoS-indexed research publication during the years 1980-2010, in some cases the quantities are very low (i.e. less than five publications on average in recent years). These are clearly not research-intensive institutes, at least not in terms of research with documented outputs in the form of research articles in scientific serial literature. Hence, in these cases the available bibliometric data were insufficient to create valid and reliable information for the bibliometric performance indicators, especially when the data is drawn from the WoS database for just a single (recent) publication year. This caveat applies to the overall profile (across all fields of science), but especially to the level of the selected fields where the quantities may become extremely low or non-existent.

In follow-up stages of U-Multirank, we plan to lower the threshold values for WoS-indexed publication output in order to discard those institutions, or fields of science, where the bibliometric indicators or measurements are no longer amenable to detailed analysis of publication output or citation impact performance. Depending on the severity of the problem within a HEI, we can then either:

- remove the institution from all indicators that involve bibliometric data;
- include bibliometric information only for the overall profile across all fields of science;

- include bibliometric information for the overall profile, as well as for those selected fields where sufficient publication output was produced in the selected time period.

The annual publication threshold for the overall profile will most likely be set at an annual average of 50-100 WoS-indexed publications, with the annual field-specific thresholds set at 10 to 15 publications.

5.3.2.2 Patent data

As indicated in chapter 4 (section 4.2.3), for our analysis of patents we collected data from the October 2009 version of the international PATSTAT-database. In this database the institutions participating in the sample were identified and studied in order to extract the institutional-level patent-data.

The extraction covers patents from the three largest patent offices worldwide: the European Patent Office (EPO), the US Patent and Trademark Office (USPTO) and the World Intellectual Property Organization (WIPO). Patent offices are selected by putting a query filter on the 'publication authority' field ('PUBLN_AUTH') in the Patent Publication Table (see appendix 7: Table TLS211_PAT_PUBLN).

The extraction of institutional-level patent data is based on identification of the institute in the applicant field of the PATSTAT database (see appendix 7: table TLS206_PERSON). The development of patent indicators on the micro-level of specific entities – such as universities – is complicated by the heterogeneity of patentee names that appear in patent documents within and across patent systems. Inconsistencies such as spelling mistakes, typographical errors and name variants (often also reflecting idiosyncrasies in the organization of intellectual property activities within organizations) considerably complicate analyses at the institutional level.

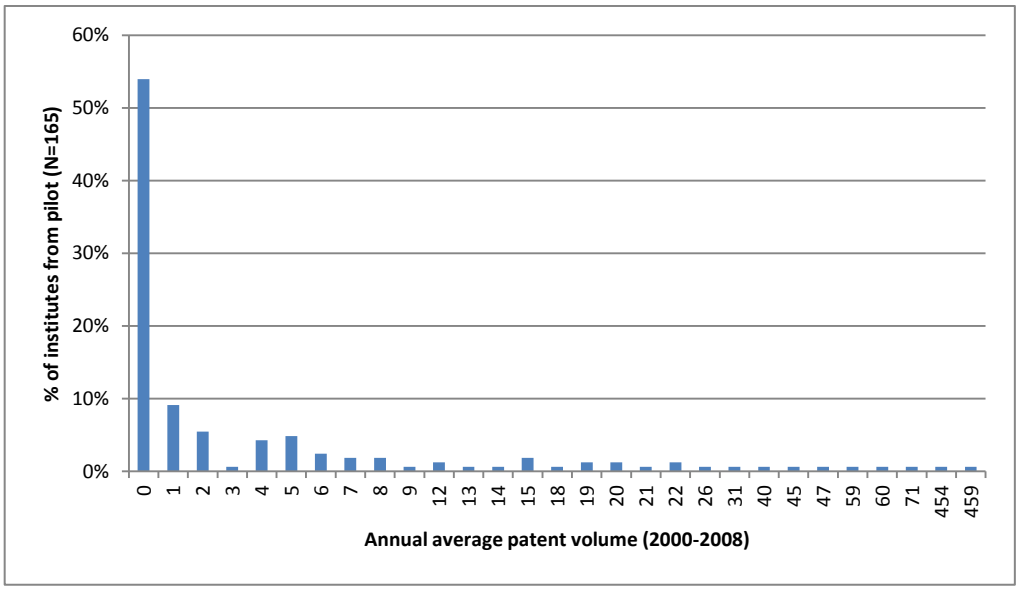
Several measures were taken to minimize the consequential chance of missing hits. First and foremost, all queries were performed on a table with *a priori* harmonized PATSTAT applicant names. The comprehensive and automated name cleaning method from which this table results, was developed by ECOOM (Centre for R&D Monitoring, Leuven University; partner in CHERPA), in partnership with Sogeti[29], in the framework of the EUROSTAT work on Harmonized Patent Statistics. Details on the name harmonization methodology can be found in the Compendium of Patent Statistics, recently published by Eurostat (2011). Second, and specifically for the U-Multirank pilot, keyword searches were designed and tailored for each institute individually, to include as many as possible known name variants. Finally, each resulting list of retrieved name variants was checked manually and, if needed, false hits were eliminated. To illustrate these institutional

[29] http://www.sogeti.com/

keyword searches, appendix 6 presents five examples of institutes, keywords, and retrieved/withheld name variants. As is the case for the bibliometric analysis (see 5.3.2.1), institutional patent information was – as yet - gathered in a 'top-down' manner, i.e. without an external 'bottom-up' verification of the extracted data by one or more representatives of each organization. As a result, although the above discussed harmonization steps imply high levels of accuracy and coverage (see also Magerman, 2009; Peeters, 2009), we cannot guarantee 100% completeness for the extracted sets of patents.

As illustrated below in Figure 5-5, patent volumes for knowledge institutes are commonly low and highly skewed. Hence, to allow for enough variation between institutional patent volumes, a sufficiently long time period was considered: all patents were counted with application years between 2000 and 2008. Even then, as can be seen in Figure 5-5, over half of the pilot institutes (N=89) had no patents in the considered 9-year period.

Figure 5-5: Distribution of annual average patent volume for pilot institutes (N = 165)



As mentioned before, the allocation of patents to institutes is based on the name of the institute that is registered as a patent applicant. Using inventor information for extracting institution-level data is impossible, as patent documents contain no (systematic) information on the institutional affiliation of individual inventors. This implies that patents, invented by one or more academic scientists, but for which the intellectual property rights are not assigned to the institution (i.e. the applicants are companies, governmental funding agencies or individual scientists), are not covered in the extraction. Several – mostly European – studies have compared the volumes of such 'university-invented' patents (invented by an academic scientist) versus 'university-owned' (with the university registered as applicant). Evidence from studies in France

(Azagra-Caro et al., 2003), Finland (Meyer et al., 2003), Belgium (Saragossi & van Pottelsberghe, 2003), Italy (Balconi et al., 2004) and Norway (Iversen et al., 2007) suggests that about 60% of university-*invented* patents are *not* university-*owned*. The available evidence from some US studies indicates much smaller percentages (approximately 20%) of university-invented patents that are not university-owned (Thursby et al., 2007).

Moreover, national and institutional differences in culture and legislation regarding intellectual property rights on university-created knowledge will cause the size of the consequential 'bias' to vary between countries. Institutional and national differences may concern the autonomy of institutions, the control they exercise over their academic staff, and the legal norms on the assignment of intellectual property rights (IPR) over academic research results. To illustrate this with an example: academic patents in Europe (i.e. patents invented by academic scientists) are much less likely to be 'owned' by universities (i.e. the university is registered as applicant) than in the USA, as European universities have lower incentives to patent or generally have less control over their scientists' activities (Lissoni et al., 2008). This does *not* mean that European academic scientists do not effectively contribute to the inventive activity taking place in their countries, as one might presume from considering only the statistics on university-owned patents. On the contrary, the data provided and discussed in the study by Lissoni et al. (2008) show that the extent of academic scientists' contribution to national patenting in France, Italy and Sweden is quite similar to that found for the USA. The difference lies in the ownership regimes: as opposed to the USA, where universities own the majority of academic patents, Europe witnesses the dominance of business companies, which own no less than 60% of academic patents. In France, and to a lesser extent in Italy, a sizeable share of academic patents is also owned by large governmental research organizations, a result which reflects the importance of these actors in their national public research systems.

As such, when interpreting institution-level patent data such as the ones provided in this study, one should at all times bear in mind the relatively sizable volume of university-invented patents that is *not* retrieved by the institution-level search strategy and institutional and national variations in the size of the consequential limitation bias.

We have argued that the field-based rankings of indicators in each dimension contribute significantly to the value and the usability of U-Multirank. At present, however, the breakdown of patent indicators by the fields defined in the U-Multirank pilot study (business studies, mechanical engineering and electrical engineering) is unfeasible, due to a lack of concordance with the field classification that is present in the patent database. The latter is organized according to the technological breakdown of the International Patent Classification. The International Patent Classification (IPC) was established by the Strasbourg Agreement 1971 and provides for a hierarchical system of symbols for the

classification of patents according to the different areas of technology to which they pertain. The IPC divides technology into eight sections with approximately 70,000 subdivisions. The IPC classes are allotted by the national or regional industrial property office that publishes the patent document. In order to keep the IPC up to date, it is continuously revised and a new version is published regularly. Hence, the classification of patents is based on technologies or products which use specific technologies. The overview of higher education fields is based on educational programs, research fields and other academically-oriented criteria. Due to the consequential large difference in notions that underlie 'higher education field' versus 'technology field', a concordance between both is meaningless. Therefore we were unable to produce patent analyses at the field-based level of U-Multirank.

6 Testing U-Multirank: results

6.1 Introduction

The main objective of the pilot study was to empirically test the feasibility of the U-Multirank instrument. In this chapter we report on the outcomes of this pilot test. We will first present the feasibility of the use of the various indicators presented in chapter 3. Next we will discuss the feasibility of the data collection procedures including the quality of the data sources. Finally we will discuss the level of institutional interest in participating in the pilot and the potential upscaling of U-Multirank to a globally applicable multidimensional ranking tool.

6.2 Feasibility of indicators

In the pilot study we analyzed the feasibility of the various indicators that were selected after the multi-stage process of stakeholder consultation. This analysis thus refers to the list of indicators presented in chapter 3.

As described in chapter 3, the selection of indicators has been based on the application of a number of criteria:

- *relevance*: the relative importance of the indicator according to the various stakeholders' perspectives
- *validity*: the indicator measures what it claims to measure. This criterion is broken down into:
    - *Concept* and *construct validity*: the indicator focuses on the performance of (programs in) higher education and research institutions and is defined in such a way that it measures 'relative' characteristics (e.g. controlling for size of the institution)
    - *Face validity:* the indicator is used in other benchmarking and/or ranking exercise and thus may be regarded as a measure of performance, which already appears to be used
- *reliability*: the measurement of the indicator is the same regardless of who collects the data or when
- *comparability*: the indicator allows comparison from one situation/system/location to another
- *feasibility*: the required data are available or can be collected with an acceptable level of effort.

Using these criteria the indicators were 'pre-selected' as the base for the pilot test. In the following tables we present both this 'pre-selection' and the results from the empirical feasibility test. For reasons of comprehension and to avoid confusion we have redefined and reordered the criteria applied in the original selection as follows:

- relevance
- concept/construct validity
- face validity
- robustness consisting of reliability and comparability
- availability (of data), instead of feasibility (because feasibility is the major subject of the pilot test).

These five criteria are presented in the left-hand columns of the tables in this section allowing a 'preliminary' (pre-pilot) rating. Rating 'A' expresses a consensus on the fitness for purpose of the indicator; rating 'B' indicates that some stakeholders and/or experts have expressed some doubts regarding one or two selection criteria. The 'relevance' criterion has been the major reason to keep these indicators on the list for the pilot study.

In the right-hand columns of the tables, the result of the empirical assessment of the feasibility of the indicators is summarized in a (post-pilot) final feasibility score. Score 'A' indicates that the feasibility is judged to be high; score 'B' indicates that there are some problems regarding the feasibility but in most cases data on the indicators can be collected and interpreted. Score 'C' indicates that there are serious problems in collecting data on the indicator.

The (post-pilot) feasibility score is based on three criteria:

- *data availability*: the relative actual existence of the data needed to build the indicator. If information on an indicator or the underlying data elements is/are missing for a relatively large number of cases, the data availability is assumed to be low.
- *conceptual clarity*: the relative consistency across individual questionnaires regarding the understanding of the indicator. If, in the information collected during the pilot study, there is a relatively large and/or diversified set of comments on the indicator in the various questionnaires, the conceptual clarity is assumed to be low.
- *data consistency*: the relative consistency regarding the actual answers in individual questionnaires to the data needs of the indicator. If in the information collected during the pilot study, there is a relatively large level of inconsistencies in the information provided in the individual questionnaires, the data consistency is assumed to be low.

Indicators which were rated 'A' or 'B' during (pre-pilot) preliminary rating but which received a 'C' in terms of the (post-pilot) feasibility score were reconsidered with regard to their inclusion in the final list of indicators. For this reconsideration process a special and final stakeholders' workshop was organized. For indicators with a problematic feasibility score there are two options:

1. They are judged highly relevant despite the problematic score and therefore efforts to enhance the data situation will be proposed; these indicators are kept 'in'.
2. They are not regarded as (very) relevant and in light of the feasibility problems they are deleted from the list of indicators ('out').

The last column (In/Out) in the tables shows the respective conclusions on those indicators based on consultation with stakeholders and the Advisory Group.

6.2.1 Teaching & Learning

The first dimension of U-Multirank is Teaching & Learning. Tables 6-1 to 6-3 provide an overview of the indicators in this dimension according to the criteria and assessments described above.

Table 6-1: Focused institutional ranking indicators: Teaching & Learning

TEACHING & LEARNING	**Rating of indicators (pre-pilot)**						**Feasibility score (post-pilot)**				
Focused institutional ranking	Relevance	Concept/construct validity	Face validity	Robustness	Availability	Preliminary rating	Feasibility score	Data availability	Conceptual clarity	Data consistency	**Recommendation**
Graduation Rate	▲	■	▲	■	■	**A**	**B**	■	▲	▲	
Time to Degree	■	■	▲	■	■	**B**	**B**	■	▲	▲	
Relative Rate of Graduate (Un)employment	▲	▲	■	■	■	**B**	**C**	■	▼	▲	**In**
Interdisciplinarity of programmes	■	■	■	▼	▼	**B**	**B**	■	■	▲	
Expenditure on teaching	■	▼	▼	▲	▼	**B**	**B**	■	■	■	

Observations from the pilot test:

- Much to our surprise there were few comments on the indicators on graduation rate and time to degree.
- Most comments were regarding graduate employment. The fact that in many countries/institutions different measurement periods (other than 18 months after graduation) are used seriously hampers the interpretation of the results on this indicator.
- A relatively high number of respondents commented that 'interdisciplinarity of programs' requires more clarification.
- The breakdown of expenditure by activity (teaching, research) appeared to be problematic in a number of institutions. For those institutions that did provide data on the breakdown, a number of institutions indicated that the estimates were rather crude.

For the field-based rankings two subsets of indicators have been distinguished: the indicators that have been built using the information from departmental questionnaires and the indicators related to student satisfaction data.

Table 6-2: Field-based ranking indicators: Teaching & Learning (departmental questionnaires)

TEACHING & LEARNING	**Rating of indicators (pre-pilot)**						**Feasibility score (post-pilot)**				
Field-based ranking Departmental questionnaire	Relevance	Concept/construct validity	Face validity	Robustness	Availability	Preliminary rating	Feasibility score	Data availability	Conceptual clarity	Data consistency	**Recommendation**
Student/staff ratio	▲	■	▲	■	▲	**A**	**A**	▲	▲	▲	
Graduation rate	▲	■	▲	■	■	**A**	**B**	■	■	▲	
Qualification of academic staff	■	▼	▼	▲	■	**B**	**A**	▲	▲	▲	
Percentage graduating within norm period	■	■	▲	■	■	**B**	**B**	■	▲	■	
Relative rate of graduate unemployment	▲	▲	■	■	▼	**B**	**C**	■	■	▼	**In**
Interdisciplinarity of programmes	▲	▼	▼	▼	■	**B**	**B**	▲	■	■	
Inclusion of work experience	■	▼	▼	■	■	**B**	**A-B**	▲	▲	■	
Gender balance	▼	■	▼	▲	▲	**B**	**A**	▲	▲	▲	

Observations from the pilot test:

- A number of institutions did not have information on graduate employment/unemployment at the field level. In addition, both institutional and national data, to which some institutions could refer, use different time periods in measuring employment status (e.g. six, 12 or 18 months after graduation). As normally the rate of employment is increasing continuously over time, particularly during the first year after graduation, comparability of data is seriously hampered by different time periods. In accordance with the institutional ranking the indicator was nevertheless regarded as highly relevant by stakeholders.
- The indicator 'inclusion of work experience' is a composite indicator using a number of data elements (e.g. internships, teachers' professional experience outside HE) on employability issues; if one of the data elements is missing, the score for the indicator cannot be calculated.

Table 6-3: Field-based ranking indicators: Teaching & Learning (student satisfaction scores)

TEACHING & LEARNING	**Rating of indicators (pre-pilot)**						**Feasibility score (post-pilot)**				
Field-based ranking Student survey	Relevance	Concept/construct validity	Face validity	Robustness	Availability	Preliminary rating	Feasibility score	Data availability	Conceptual clarity	Data consistency	**Recommendation**
Organization of programme	▲	▲	■	■	▲	**A**	**A**	▲	▲	▲	
Inclusion of work experience	▲	▲	■	■	▲	**A**	**A**	▲	▲	▲	
Evaluation of teaching	▲	▲	■	■	▲	**A**	**A**	▲	▲	▲	
Social climate	▲	▲	■	■	▲	**A**	**A**	▲	▲	▲	
Quality of courses	▲	▲	■	■	▲	**A**	**A**	▲	▲	▲	
Support by teacher	▲	▲	■	■	▲	**A**	**A**	▲	▲	▲	
Computer facilities	▲	▲	■	■	▲	**A**	**A**	▲	▲	▲	
Overall judgment	▲	▲	■	■	▲	**A**	**A**	▲	▲	▲	
Libraries		▲	■	■	▲	**B**	**A**	▲	▲	▲	
Laboratories		▲	■	■	▲	**B**	**A**	▲	▲	▲	

There are no major problems with regard to the feasibility of individual indicators from the student survey. General aspects of feasibility of a global student survey are discussed in section 6.3.

6.2.2 Research

Indicators on research include bibliometric indicators (institutional and field-based) as well as indicators derived from institutional and field-based surveys. In general the feasibility of the research indicators, which are the main focus of existing international rakings, is judged to be good; nevertheless some indicators turned out be problematic.

Table 6-4: Focused institutional ranking indicators: Research

RESEARCH	**Rating of indicators (pre-pilot)**						**Feasibility score (post-pilot)**				
Focused institutional ranking	Relevance	Concept/construct validity	Face validity	Robustness	Availability	Preliminary rating	Feasibility score	Data availability	Conceptual clarity	Data consistency	**Recommendation**
Percentage of expenditure on research	▲	▼	■	▼	■	**A**	**B**	■	■	▲	
Field-normalized citation rate *	▲	▲	■	▲	▲	**A**	**A**	▲	▲	▲	
Post-docs per fte academic staff	▲	▲	■	▲	■	**A**	**B**	■	■	▲	
Percentage research income from competitive sources	■	▲	▼	■	■	**B**	**B**	■	▲	▲	
Art-related outputs per fte academic staff	▼	■	▼	▼	▼	**B**	**C**	▼	▼	▲	**In**
Total publication output	▼	■	▼	■	▲	**B**	**B**	■	■	▲	
International awards and prizes won	▲	■	▼	▼	▼	**B**	**C**	▼	■	▲	**Out**
Highly cited research publications *	▲	■	■	▲	▲	**B**	**A**	▲	▲	▲	
Interdisciplinary research activities	▲	■	▼	▲	■	**B**	**A**	▲	▲	▲	

* Data source: bibliometric analysis

Observations from the pilot test:

- The comments regarding expenditure on research refer to the problem of breaking down the basic government funding provided as a lump sum.
- The comments on the 'post-doc' positions mainly regarded the clarity of definition and the lack of proper data.
- The large number of missing data and comments regarding the art-related output was no surprise. The lack of clarity in the definition corroborated the high number of missing values in this indicator. Stakeholders, in particular representatives of art schools, stressed the relevance of this indicator despite the poor data situation. The neglect of research performance in the arts and art-related fields is a major flaw of existing rankings. Even if this deficit cannot be overcome immediately, efforts should be made to enhance the data situation on cultural research outputs of higher education institutions. This cannot be done by

producers of rankings alone; initiatives should also come from providers of (bibliometric) databases as well as stakeholder associations in the sector.

Table 6-5: Field-based ranking indicators: Research

RESEARCH	**Rating of indicators (pre-pilot)**						**Feasibility score (post-pilot)**				
Field-based ranking	Relevance	Concept/construct validity	Face validity	Robustness	Availability	Preliminary rating	Feasibility score	Data availability	Conceptual clarity	Data consistency	**Recommendation**
External research income	▲	▲	■	■	▲	**A**	**A**	▲	■	▲	
Total publication output *	▲	▲	■	■	■	**A**	**A**	▲	▲	▲	
Student satisfaction: research orientation of program	■	■	▼	■	▲	**A**	**A**	▲	▲	▲	
Doctorate productivity	▼	▲	▼	▲	▲	**B**	**A**	▲	▲	▲	
Field-normalized citation rate *	▲	▲	■	▲	▲	**B**	**A**	▲	▲	▲	
Highly cited research publications*	▲	■	■	▲	▲	**B**	**A**	▲	▲	▲	
Post-docs per PhD completed	*New indicator*						**B**	■	■	▲	

* Data source: bibliometric analysis

Observations from the pilot test:

- On the field level, the proposed indicators do not encounter any major feasibility problems. In general, the data delivered by faculties/departments revealed some problems in clarity of definition of staff data. In particular the understanding and handling of the concept of 'full-time equivalents' (fte), which is used as a reference point to standardize indicators for size effects, proved difficult. Here a clearer yet concise explanation (including an example) should be used in future data collection.
- It was also noted that the relevance and the exactness of definition of 'post-doc' positions differ across fields. The data on post-doc decisions proved to be more problematic in business studies than in engineering. With regard to future applications in other fields this must be kept in mind: while post-doc positions are very common in the sciences they are less widespread in the social sciences and not clearly defined in humanities.

6.2.3 *Knowledge transfer*

The dimension of knowledge transfer is, together with the regional engagement dimension, almost completed neglected in existing rankings, both nationally and internationally.

Table 6-6: Focused institutional ranking indicators: Knowledge Transfer

KNOWLEDGE TRANSFER	**Rating of indicators (pre-pilot)**						**Feasibility score (post-pilot)**				
Focused institutional ranking	Relevance	Concept/construct validity	Face validity	Robustness	Availability	Preliminary rating	Feasibility score	Data availability	Conceptual clarity	Data consistency	**Recommendation**
Percentage of income from third party funding	▲	■	▼	▼	■	**A**	**C**	▼	▲	■	**In**
Incentives for knowledge transfer	▲	■	▼	▼	■	**A**	**A**	▲	▲	▲	
Patents awarded**	■	▲	▲	▲	▲	**A**	**B**	■	▲	▲	
University-industry joint research publications *	▲	▲	■	▲	▲	**A**	**A**	▲	▲	▲	
CPD courses offered per fte academic staff	▲	■	▼	▼	▼	**B**	**B**	■	▲	▲	
Start-ups per fte academic staff	■	▼	■	■	■	**B**	**B**	■	▲	▲	
Technology transfer office staff per fte academic staff	■	■	■	■	▲	**B**	**B**	■	■	▲	
Co-patenting **	■	▲	■	▲	▲	**B**	**A**	▲	▲	▲	

* Data source: bibliometric analysis; ** patent analysis

Observations from the pilot test:

- The indicators related to knowledge transfer did not cause much comment. Comments on TTO staff were mainly on the different way technology transfer activities are organized at the institutional level, making it difficult to compare the data.

Table 6-7: Field-based ranking indicators: Knowledge Transfer

KNOWLEDGE TRANSFER	**Rating of indicators (pre-pilot)**						**Feasibility score (post-pilot)**				
Field-based ranking	Relevance	Concept/construct validity	Face validity	Robustness	Availability	Preliminary rating	Feasibility score	Data availability	Conceptual clarity	Data consistency	**Recommendation**
University-industry joint research publications *	▲	▲	▼	▲	▲	**A**	**A**	▲	▲	▲	
Academic staff with work experience outside HE	▲	▲	▼	■	■	**A**	**B**	■	■	■	
Joint research contracts with private enterprise	▲	▲	▼	■	▲	**A**	**B**	■	▲	▲	
Patents awarded **	▼	▲	▲	▲	▼	**C**	**C**	▼	▲	▼	**Out**
Co-patenting **	▼	▲	▼	▲	▲	**B**	**C**	▼	▲	▼	**Out**
Annual income from licensing	▼	▲	▼	■	▼	**B**	**C**	▼	■	■	**Out**
Number of licensing agreements	▲	▲	▼	■	▼	**B**	**C**	▼	■	■	**Out**

* Data source: bibliometric analysis; ** patent analysis

Observations from the pilot test:

- In contrast to the findings at institutional level, the feasibility of the Knowledge Transfer indicators turned out to be highly problematic for field-based rankings. The only indicator with an 'A'-rating – indicating a high degree of feasibility – comes from bibliometric analysis.
- Availability of data on 'joint research contracts with private sector' is a major problem, but primarily in business studies and less in engineering.
- The indicators based on data from patent databases are feasible only for institutional ranking due to discrepancies in the definition and delineation of fields in the databases.
- Only a small number of institutions could deliver data on licensing.
- There was an agreement among stakeholders, therefore, that those indicators should be used for focused institutional rankings only.

6.2.4 *International orientation*

Most of the indicators on the dimension 'international orientation' proved to be relatively unproblematic in terms of feasibility.

Table 6-8: Focused institutional ranking indicators: International Orientation

INTERNATIONAL ORIENTAION	**Rating of indicators (pre-pilot)**						**Feasibility score (post-pilot)**				
Focused institutional ranking	Relevance	Concept/construct validity	Face validity	Robustness	Availability	Preliminary rating	Feasibility score	Data availability	Conceptual clarity	Data consistency	**Recommendation**
Percentage of programs in foreign language	▲	■	■	■	▲	**A**	**A**	▲	▲	▲	
International joint research publications*	▲	▲	■	▲	▲	**A**	**A**	▲	▲	▲	
Percentage of international staff	▲	■	■	■	■	**B**	**A**	▲	▲	▲	
Percentage of students in international joint degree	▲	▲	■	■	▲	**A**	**B**	■	■	▲	
International doctorate graduation rate	▼	■	▼	■	■	**B**	**A**	▲	▲	▲	
Percentage foreign degree-seeking students	*New indicator*						**B**	■	▲	▲	
Percentage students coming in on exchanges	*New indicator*						**A**	▲	▲	▲	
Percentage students sent out on exchanges	*New indicator*						**A**	▲	▲	▲	

* Data source: Bibliometric analysis

Observations from the pilot test:

- There were some problems reported with availability of information on nationality of qualifying diploma and students in international joint degree programs. In the latter, problems related primarily to the inaccuracy of the definition and the problems in interpretation stemming from this.

Table 6-9: Field-based ranking indicators: International Orientation

INTERNATIONAL ORIENTAION	**Rating of indicators (pre-pilot)**						**Feasibility score (post-pilot)**				
Field-based ranking	Relevance	Concept/construct validity	Face validity	Robustness	Availability	Preliminary rating	Feasibility score	Data availability	Conceptual clarity	Data consistency	**Recommendation**
Percentage of international students	▪	▲	▪	▪	▲	**A**	**A**	▲	▲	▲	
Incoming and outgoing students	▲	▲	▪	▪	▲	**A**	**A-B**	▪	▪	▲	
Opportunities to study abroad (student satisfaction)	▪	▪	▼	▪	▲	**A**	**B**	▪	▲	▲	
International orientation of programs	▲	▲	▼	▪	▪	**A**	**B**	▪	▲	▲	
International academic staff	▲	▪	▪	▪	▪	**B**	**A-B**	▪	▲	▪	
International joint research publications*	▲	▪	▼	▲	▲	**B**	**A**	▲	▲	▲	
International research grants	▼	▪	▪	▪	▲	**B**	**B**	▪	▲	▲	
International doctorate graduation rate	▲	▪	▼	▪	▪	**B**	**A**	▲	▲	▲	

* Data source: Bibliometric analysis

Observations from the pilot test:

- Not all institutions have clear data on outgoing students. In some cases only those students participating in institutional or broader formal programs (e.g. ERASMUS) are registered and institutions do not record numbers of students with self-organized stays at foreign universities.
- Availability of data was relatively low regarding the student satisfaction indicator as only a few students had already participated in a stay abroad and could assess the support provided by their university.
- The indicator 'international orientation of programs' is a composite indicator referring to several data elements; feasibility is limited by missing cases for some of the data elements.
- Some institutions could not identify external research funds from international funding organizations.
- In order to test alternatives means of measuring percentages of international staff, we used different definitions in the institutional and field-based rankings. The

institutional questionnaire referred to the nationality of staff; the level of staff with foreign nationality was easy to identify for most institutions. In the field questionnaires, the definition 'international' referred to staff hired from abroad. This excludes foreign staff who were hired from another institution in the same country rather than from abroad. Some universities had difficulties to identify their international staff based on this definition.

6.2.5 *Regional engagement*

Up to now the regional engagement role of universities has not been included in rankings. There are a number of studies on the regional economic impact of higher education and research institutions, either for individual institutions and their regions or on higher education in general. Those studies do not offer comparable institutional indicators or indicators disaggregated by fields.

Table 6-10: Focused institutional ranking indicators: Regional Engagement

REGIONAL ENGAGEMENT	**Rating of indicators (pre-pilot)**						**Feasibility score (post-pilot)**				
Focused institutional ranking	Relevance	Concept/construct validity	Face validity	Robustness	Availability	Preliminary rating	Feasibility score	Data availability	Conceptual clarity	Data consistency	**Recommendation**
Percentage of income from regional sources	▲	■	■	■	▲	**A**	**C**	▼	■	▲	**In**
Percentage of graduates working in the region	▲	▲	▼	▼	▼	**B**	**C**	▼	■	▼	**In**
Research contracts with regional partners	▲	■	▼	▼	■	**B**	**B**	■	▲	▲	
Regional joint research publications *	▼	■	▼	▲	▲	**B**	**A**	▲	▲	▲	
Percentage of students in internships in local enterprises	■	▲	▼	▼	▼	**B**	**C**	▼	■	▲	**In**

* Data source: Bibliometric analysis

Observations from the pilot test

- A general comment regarding the indicators of regional engagement on both institutional and field level related to the delineation of the region. The NUTS regions are not applicable outside Europe, which caused some problems in non-European higher education institutions. But even within Europe NUTS regions are seen as problematic by some institutions, in particular those from smaller

countries having only one or two NUTS 2 regions. Although the conceptual clarity on the issue is good , the low level of data consistency showed that there is a wide variety of region definitions used by institutions, which may harm the interpretation of the related indicators.

- Both in institutional and in field-based data collection information on regional labor market entry of graduates could not be delivered by most institutions. Here the problems concerning the availability of comparable information on graduate employment in general and the problems with the definition/delineation of `region' add up. There is a clear perception of the relevance of employability issues, and the relevance of higher education and research to the regional economy and the regional society at large, and stakeholders were strongly in favor of keeping the indicator (both for institutional and for field-based rankings).
- The most feasible indicator is the bibliometric indicator `Regional co-publications'. Here region can be defined either by NUTS regions or in a more flexible way by the distance between locations of the collaborating institutions.

Table 6-11: Field-based ranking indicators: Regional Engagement

REGIONAL ENGAGEMENT	**Rating of indicators (pre-pilot)**						**Feasibility score (post-pilot)**				
Field-based ranking Departmental questionnaire	Relevance	Concept/construct validity	Face validity	Robustness	Availability	Preliminary rating	Feasibility score	Data availability	Conceptual clarity	Data consistency	**Recommendation**
Graduates working in the region	▲	▲	▼	▼	▼	**B**	**C**	▼	■	▼	**In**
Regional participation in continuing education	▲	■	▼	▼	▼	**B**	**C**	▼	■	■	**Out**
Student internships in local enterprises	▲	■	▼	▼	■	**B**	**B-C**	■	■	▲	**In**
Degree theses in cooperation with regional	■	■	▼	■	■	**B**	**B-C**	■	■	▲	**In**
Summer schools	▼	▼	▼	▼	■	**C**	**C**	▼	■	▲	**Out**
Regional joint research publications *	*New indicator*						**A**	▲	▲	▲	

* Data source: bibliometric analysis

Observations from the pilot test:

- Less than half of the pilot institutions could deliver data on regional participation in continuing education programs (and only one fifth in mechanical engineering programs). Based on feedback from institutions and stakeholders, this indicator cannot be seen as feasible; there is probably no way to improve the data situation in the short term.
- While far from good, the data situation on student internships in local enterprises and degree theses in cooperation with local enterprises turned out to be less problematic in business studies than that found in the engineering field. Both internships and degree theses enable the expertise and knowledge of local higher education institutions to be utilized in a regional context, in particular in small- and medium-sized enterprises. At the same time they are a link to potential future employees and in many non-metropolitan regions they play an important role in the recruitment of higher education graduates.

6.3 Feasibility of data collection

As explained in section 5.3 data collection during the pilot study was carried out via self-reporting from the institutions and analysis of international bibliometric and patent databases.

6.3.1 Self-reported institutional data

For the collection of self-reported institutional data we made use of several questionnaires:

- the U-map questionnaire to identify institutional profiles
- the U-Multirank institutional questionnaire
- the U-Multirank field-based questionnaire

We supported this data collection with extensive data cleaning processes, plausibility checks and feasibility loops with the institutions participating in the sample. In addition we undertook a follow-up survey in order to further assess the feasibility of the data collection.

In general the organization and procedures of the self-reported institutional data collection were evaluated as largely positive or at least 'neutral' by the institutions. Very few institutions were really dissatisfied with the processes. The collection of data by online questionnaires worked well, and the coordination of all data collection via a central contact person in participating institutions also proved successful.

We made the following key observations regarding the process of collecting self-reported institutional data:

- The parallel institutional data collection for U-Map and U-Multirank caused some confusion. Although a tool was implemented to pre-fill data from U-Map into U-Multirank, some confusion remained concerning the link between the two instruments.
- In order to test some varieties, institutional and field-based questionnaires were implemented with different features (e.g. definition of international staff). This procedure helped us to judge the relative feasibility of concepts and procedures.
- The glossary of indicators and data elements proved helpful in achieving a high degree of consistency in the data delivered by the institutions. Yet the definitions and explanations of some elements (e.g. staff categories including fte, the delineation of regions) could be improved – bearing in mind that there is an apparent trade-off between adequate explanation and willingness to read a lengthier explanation. One option might be to build the glossary into the online questionnaire so that respondents have the explanations at hand while supplying their data.
- The effort to include a feedback cycle both in institutional and field-based data collection (with questions and comments on the data already submitted) was greatly appreciated by the institutions. Although it implied a major investment of time by the project team, this procedure proved to be very efficient and helped significantly to increase the quality and consistency of the data.
- In some countries the U-Multirank student survey conflicted with existing national surveys, which in some cases are highly relevant for institutions. It should be evaluated how far U-Multirank and national surveys could be harmonized in terms of questionnaires and, at least, in terms of timing.
- While a field period of four to six weeks after sending out invitations to students seems appropriate at individual institutions, the time window to organize a student survey across all institutions has to be at least six months in order to avoid conflicts with the various academic calendars (vacations and examination periods differ substantially between countries).

Our major conclusion regarding the feasibility of the self-reported institutional data is that data availability is an issue in a number of cases. This is not so much a problem of the multidimensional ranking tool itself, but with the administrative processes related to data collection in some institutions. It may be assumed that when institutions increase their efforts regarding data collection and data quality this problem will be mitigated.

6.3.2 Student survey data

One of the major challenges regarding the feasibility of our global student survey is whether the subjective evaluation of their own institution by students can be compared globally or whether there are differences in the levels of expectations or respondent behavior. Survey research among different groups of respondents with different national and cultural background must take into account that the respondents may have different standards by which they evaluate situations or events.

In our student questionnaire we used 'anchoring vignettes' to control for such effects. Anchoring vignettes is a technique designed to ameliorate problems that occur when different groups of respondents understand and use ordinal response categories to evaluate services and social situations in general (cf. King et al 2004, King and Wand 2006). Anchoring vignettes make it possible to construct a common scale of measurement across respondent groups by constructing a hypothetical situation which is assessed by these respondents. Anchoring vignettes have been tested and used e.g. in health service research; up to now they have not been used in comparative higher education research. Hence we had to develop our own approach to this research technique. (For a detailed description see appendix 9)

Our general conclusion from the anchoring vignettes analysis was that no correlation could be found between the students' evaluation of the situation in their own institutions and the expectation levels as reflected in our anchoring vignettes. This implies that the student assessments were not systematically influenced by differences in levels of expectation (related to different national backgrounds or cultures), and thus that the feasibility of the data collection through a global-level student survey is sufficiently feasible.

6.3.3 Bibliometric and patent data

The collection of bibliometric and patent data turned out to be largely unproblematic. However, a few observations need to be made.

As indicated in chapter 5, in bibliometric analysis the sets of publications produced by a specific institution (or a subunit of it) have to be identified in international bibliographic databases. To assign the publications to a given institution, a 'top-down' delineation exercise is performed, in which the institution is automatically detected by lexical queries on the author's affiliation field (the address field) of the publications in the databases, by a query on keywords. The query is based on more or less sophisticated name patterns. But it is quite difficult to achieve a satisfactory detection since authors can use a large number of variants for the name of a given institution, or even not name it at all. The variants arise because many institutions have both current and old versions of their

names, which can be written in full or abbreviated – in addition to the unavoidable misspellings and other transcription errors. The absence of any explicit reference to the institution is often due to the exclusive use of the name of a subunit of the institution (department, institute, unit, network, lab or even team) or of a coordinating body such as a network or consortium.

To overcome these common problems, professional teams specialized in bibliometrics have developed sophisticated pre-processing methods for cleaning and normalizing the addresses before using them to identify institutions by a top-down approach. But even in such favorable cases – not always encountered in the current ranking exercises – it has been recommended that the top-down approach be complemented , as far as possible, by a bottom-up one. In a bottom-up process, representatives of the institutions are asked to identify the addresses of their publications. However, this bottom-up approach is obviously more costly and time-consuming.

In our pilot study we limited ourselves to the top-down approach only (similar to all existing international rankings). As suggested in chapter 5 (5.3.2.1)therefore, completeness of the selected bibliometric data cannot be fully guaranteed.

To assess the feasibility of our bibliometric data collection we studied the potential effects of a bottom-up verification process via a special case study of six French universities. The aim of the case study was to shed light on how a bottom-up verification approach might collect relevant data that would otherwise be missed. The case study showed that in some cases a substantial number of the publications might have been missed if relying solely on the top-down approach. However, it should be pointed out that these problematic cases occur in French institutions with particularly complex organizational structures. The statistical findings are unlikely to be representative for the institutions in our sample. The case study is reported in more detail in appendix 8.

Furthermore, several additional problems arise when the institutional delineation is to be done at a global level. First, the numerous differences between national research systems worldwide make it difficult to design global methods for defining and enriching the queries. For example, even a seemingly universal name such as ‘university’ may not describe the same institutional reality in different systems – in England or in the US, some so-called ‘universities’ could be in fact umbrella organizations covering several autonomous universities, while in France many universities are thematic and issue from one comprehensive ‘root’ university. Second, most of the national research systems tend to become more complex under the pressure of the ‘funding on project’ policies that induce the setup of various network-like institutions such as consortia, platforms and ‘poles’. Since these networks are largely built by coordination between sub-organizations, they tend to blur the borders of the institutions. For communication purposes, authors may prefer to replace the name of the university with the name of a

network. In addition, in some countries like France, universities, schools and research institutions can be interwoven by many joint labs, making it even more difficult to distinguish the institutional borders. Third, since most of the research systems worldwide are rapidly evolving, new institutions – and names – are being created all the time.

Nevertheless, the feasibility of the bibliometric data collection in the pilot study can be judged to be high. Data were easily identified and analyzed, although a warning against placing too much dependence on the completeness of the data remains in place.

With respect to the collection of patent data (via PATSTAT) there are two important caveats.

First, as mentioned before, we were only able to identify our sample **institutions** in the database. Subunits for field analyses could not be found. This implies that patents for which the intellectual property rights are assigned to companies, governmental funding agencies or individual scientists are not retrieved because the institution's name does not appear in the applicant field. Several studies have shown that the volume of such university-invented patents is sizable (Azagra Caro et al., 2003; Balconi et Potterie, 2003; Schmiemann and Durvy, 2003).

A second important caveat when extracting institutional-level patent data is that organizations register their patents under many different names and spelling variations. PATSTAT data are no exception: applicant names are often misspelled, and their spelling varies from one patent to the other. To minimize the consequential chance of missing hits we performed special keywords searches to include as many as possible name variants to the best of our knowledge. In addition, we relied on a name harmonization methodology for capturing as many name variants as possible (for details on this methodology: see Magerman et al., 2009; Peeters et al., 2009).

Also with respect to the collection of patent data the conclusion regarding its feasibility is positive. However, it should be noted that patent data analysis could only be undertaken at the institutional and not field level.

6.4 Feasibility of up-scaling

The pilot test included a limited number of institutions and only two fields. An important feasibility issue is up-scaling: is it possible to extend U-Multirank to a comprehensive global coverage and how easy would it be to add additional fields?

In terms of the feasibility of U-Multirank as a potential new global ranking tool, the results of the pilot study are positive, but with one important caveat.

The level of institutional interest in participating in the new transparency tool was encouraging. In broad terms, half of the institutions invited to participate in the pilot study agreed to do so. Given that a significant number of these institutions (32%) were from outside Europe, and taking into account that it is clear that U-Multirank is a Europe-based initiative, this represents a strong expression of worldwide interest.

However, it is important to recognize that a pilot study is not a real ranking. The institutions participating in the pilot project have access to the institutional performance profiles of all the institutions in the pilot, as well as the dimension and indicator outcomes. While this provides a unique opportunity to compare and benchmark with over 100 other institutions worldwide, the outcomes of the pilot rankings will not be made public. The overall objective of the pilot study was to design a multidimensional ranking tool and to test the feasibility of this instrument, not to publish a ranking. We may assume that the interest in a real multidimensional ranking will be substantially greater.

Our single caveat concerns an immediate global-level introduction of U-Multirank. The pilot study suggests that a global multidimensional ranking is unlikely to prove feasible in the sense of achieving extensive coverage levels across the globe in the short term. It proved particularly difficult to recruit institutions from the USA and China for the pilot project. Higher education and research institutions in the USA showed very limited interest in the study, while in China formal conditions appeared to hamper the participation of institutions. On the other hand, institutions in Australia and in a number of developing countries, largely invisible in existing global rankings, were enthusiastic about the project.

The prospects for widespread European coverage are encouraging. A substantial number of institutions both from EU and non-EU European countries participated in the projects. From their participation in the various stakeholder meetings, we can conclude that there is broad interest in the further development and implementation of U-Multirank.

We also expect that there will be continuing interest from outside Europe from institutions wishing to benchmark themselves against European institutions. And we believe that there are opportunities for the targeted recruitment of groups of institutions from outside Europe of particular interest to European higher education.

A final aspect of feasibility in terms of institutional participation is the question of institutional drop-out and non-completion rates. A brief survey of the institutions that agreed to participate but at the end of the day did not submit data suggests that data (non-) availability was a common theme. One particular group of institutions (LERU) unfortunately took a policy decision to withdraw from the project during the pre-test phase. It seems that there was a misunderstanding about 1) the function of the pre-test in

the project and 2) the conceptual interpretation of the 'user-driven approach' applied in U-Multirank. It would be interesting to involve LERU again during a follow-up project which would be to their benefit in ensuring that the selection of indicators reflects the activities of research intensive institutions. Beyond these two factors a diverse range of particular institutional issues came into play – including competing claims on the time of the staff concerned and changes in these key staff. Nevertheless, for a pilot study a completion rate of 109 of 159 (69%) is more than respectable.

The other aspect of the potential up-scaling of U-Multirank is the extension to other fields. Any extension of U-Multirank to new fields must deal with two questions:

- the relevance and meaningfulness of existing indicators for those fields, and,
- the identification and development of new field-specific indicators.

While the U-Multirank feasibility study focused on the pilot fields of business studies and engineering, some issues of up-scaling to other fields have been discussed in the course of the stakeholder consultation. Experience from the CHE ranking and other field-based rankings show that there is a core set of indicators that is relevant and meaningful for (virtually) all fields.

However, these issues do not concern all dimensions in the same way. While students can be asked about their learning experience in the same way across different fields (although questions should refer to field-specific aspects as e.g. quality of laboratory courses in technical and experimental fields) and while internationalization can be measured in similar ways across fields, other dimensions will need to have indicators that are adapted to the field concerned. A well-known example is the difference between publication cultures in the sciences/medicine and those in the humanities/social sciences. (cf. van Raan 2003) which will require different definitions of indicators of research output across different disciplinary fields. Similarly some disciplines may see dimensions such as knowledge transfer or regional engagement as less relevant to their core activities.

Any extension to additional fields therefore has to address the issue of additional specific indicators relevant to those fields. In medicine, for instance, specific indicators referring to bedside teaching and clinical education are relevant indicators in the teaching and learning dimension. Following the user- and stakeholder-driven approach of U-Multirank, we suggest that field-specific indicators for international rankings should be developed together with stakeholders from these fields. We encourage stakeholders and organizations to actively participate in the development of relevant field-specific indicators, in particular in those areas and fields which so far have largely been neglected in international rankings due to the lack of adequate data and indicators.

In the two pilot fields of business studies and engineering we were able to use 86% of the final set of indicators in both fields. We expect that when additional fields are addressed in U-Multirank, some specific field indicators will have to be developed. Based on the experience of the CHE ranking this will vary by field with some fields requiring no additional indicators and other specialized fields (such as medicine) needing up to 30% of the indicators to be tailor-made.

In general terms, we conclude that up-scaling in terms of addressing a larger number of fields in U-Multirank is certainly feasible.

7 Applying U-Multirank: presenting the results

7.1 Introduction

The quality of a ranking to a large extent depends on the quality and user-friendliness of the presentation of its results. In the past, rankings were mainly published in static print form, but for a number of years many rankings have opted for online publication (replacing or in addition to print publication). In most rankings the tables can now be sorted by individual indicators as a minimum degree of interactivity. A few rankings (e.g. the Taiwanese College Navigator published by HEEACT[30] and CHE ranking) implemented tools to produce a personalised ranking, based on user preferences and priorities with regard to the set of indicators. This approach implies the user-driven notion of ranking which also is a basic feature of U-Multirank.

The presentation of U-Multirank results outlined in this chapter strictly follows this user-driven approach. But by relating institutional profiles (created in U-Map) with multidimensional rankings, U-Multirank introduces a second level of interactive ranking beyond the user-driven selection of indicators: the selection of a sample of institutions to be compared in focused rankings. Existing international rankings are largely limited to one 'type' of institution only: internationally-oriented research universities. U-Multirank has a much broader scope and intends to include a wider variety of institutional profiles. We argue that it does not make much sense to compare institutions across diverse institutional profiles. Hence U-Multirank offers a tool to identify and select institutions that are truly comparable in terms of their institutional profiles.

7.2 Mapping diversity: combining U-Map and U-Multirank

From the beginning of the U-Multirank project one of the basic aims was that U-Multirank should be – in contrast to existing global rankings which brought about a dysfunctional short-sightedness on 'world-class research universities' – a tool to create transparency regarding the diversity of higher education institutions. The bias of existing rankings towards one specific institutional profile appears to result in the devaluing of other institutional profiles and decreasing diversity in higher education systems (see chapter 1).

Our pilot sample includes institutions with quite diverse missions, structures and institutional profiles. We have applied the U-Map profiling tool to specify these profiles.

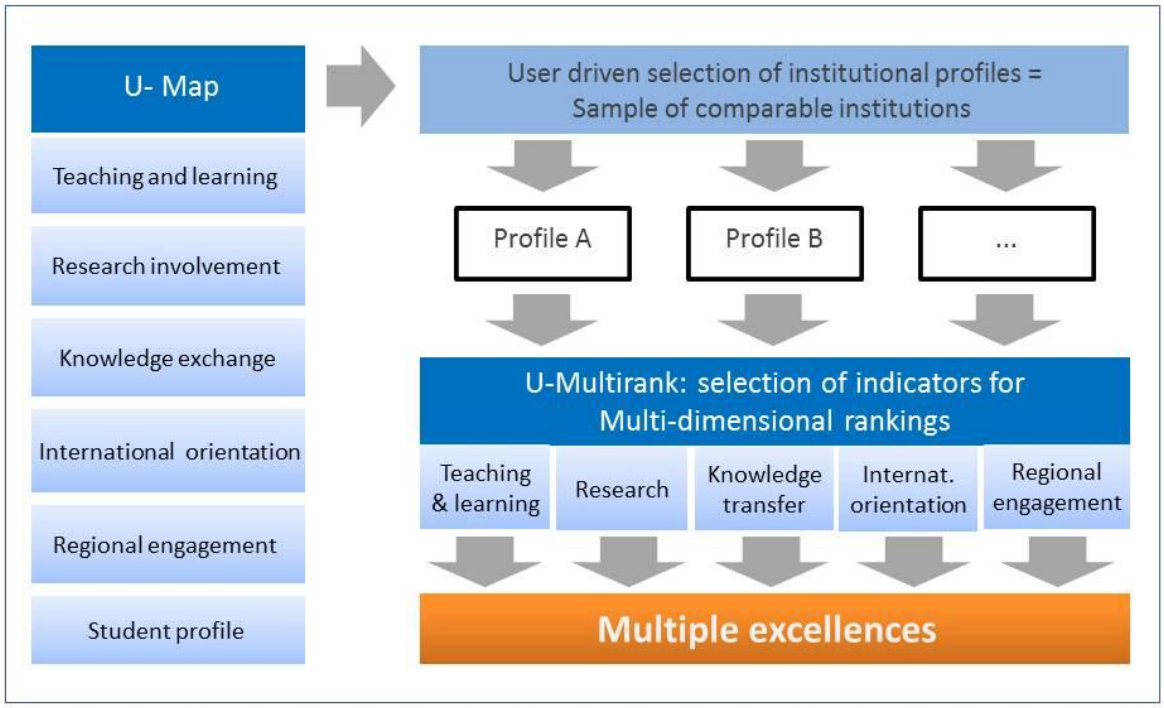
---

[30] College Navigator: http://cnt.heeact.edu.tw/site1/index2.asp?method=eintro ; CHE ranking: http://ranking.zeit.de/che2011/en/ (both retrieved on 9 May 2011)

U-Map offers a multidimensional description of profiles in six dimensions. It is user-driven in the sense that there are no fixed categories or types of institutions. Instead, users can create their own profiles by selecting indicators relevant to them out of the six dimensions.

The combination of U-Map and U-Multirank offers a new approach to user-driven rankings. Users can not only select performance indicators according to their own preferences and priorities; they can also define the institutional profile they are interested in and hence the sample of institutions to be compared in U-Multirank.

Figure 7-1: Combining U-Map and U-Multirank



Our user-driven interactive web tool will imply both steps, too. Users will be offered the option to decide if they want to produce a focused institutional ranking or a field-based ranking, and in the latter case they can select the field(s). The next step will then be the selection of the institutional profile the user is interested in. This selection defines the sample of institutions that will be included in the ranking. The user will have the option of selecting criteria from all U-Map dimensions or focusing on a specified set of dimensions. In a third step the user selects the ways the results will be presented. U-Multirank includes different ways of presenting the results.

7.3 The presentation modes

Presenting ranking results requires a general model for accessing the results, including provision for guiding users through the data and a visual framework to display the result data. In U-Multirank the presentation of data allows for both:

- a comparative overview on indicators across institutions, and,
- a detailed view of institutional profiles.

The ideas presented below are mainly inspired by the U-Map visualisations and the presentation of results in the CHE ranking.

U-Multirank produces indicators and results on different levels of aggregation leading to a hierarchical data model:

- Data at the level of institutions (results of focused institutional rankings)
- Data at the level of departments (results of field-based rankings)
- Data at the level of programs (results of field-based rankings)

The presentation format for ranking results should be consistent across the three levels while still accommodating the particular data structures on those levels.

We suggest the following modes of presentation: interactive overview (7.3.1.), personalised ranking tables (7.3.2), institutional results at a glance (7.3.3) and a detailed listing of results for single institutions, departments and programs (7.3.4.).

7.3.1 Interactive tables

The most common format used in ranking results is a table listing all institutions included in the ranking and all (or a selection of) indicators. In league table rankings tables are usually sorted by rank position. In U-Multirank we present the results alphabetically or by rank groups (see chapter 2).

In the first layer of the table (field-based ranking), an overview is presented comprising three selected indicators per dimension, a total of 15 indicators. The table displays the ranking groups (in different colours) representing the relative scores on the indicators. The current table is a 'default' table. The selection of the indicators in this table will eventually be user-driven. Based on the actual choices made by users in formulating their personalised ranking tables (see section 7.3.2) the indicators chosen most frequently will be presented in the default table.

Table 7-1: Default table with three indicators per dimension

	Teaching & Learning			Research			Knowledge transfer			international orientation			Regional engagement		
	student staff ratio	graduation rate	qualification of academic staff	research publication output	external research income	citation index	% income third party funding	CPD courses offered	startup firms	international academic staff	% international students	joint international publ.	graduates working in the region	student internships in local enterprise	regional co-publication
Institution 1	🟡	-	-	🟡	🟢	🟢	🟢	🟡	🟢	🟡	🟡	🟢	🟢	🟡	🟢
Institution 2	🔴	🟡	🟡	🟢	🟢	🟢	-	🔴	🔴	🟡	🔴	🟡	🔴	🔴	🔴
Institution 3	🟡	🟢	🟢	🟡	🟢	🟡	🟡	🟢	🟡	🟡	🟢	🟡	🟡	-	🟡
Institution 4	🟢	🟡	🟡	🟢	🔴	🟡	🔴	🟡	🟡	🟢	🟡	🔴	🟢	🔴	🟡
Institution 5	🟡	🟢	🟢	🔴	🟢	🟢	🟡	🟢	🔴	🟡	🔴	🟢	🟡	🟢	-
Institution 6	🔴	🔴	🟡	🔴	-	🔴	🟢	🟡	🟢	🔴	🟢	🟢	🟢	🟡	🟢
Institution 7	🟡	🟡	🔴	🟡	-	🔴	🟢	-	🟢	-	🟡	🟡	🟡	🔴	🟡
Institution 8	🟢	🟡	🟡	🟡	-	🟡	🟡	🔴	🟡	🔴	🟡	-	🟡	🟢	🟢
Institution 9	🟡	🔴	🔴	🟡	🔴	🔴	🔴	🟡	🔴	🟢	🟡	🔴	🔴	🟡	🔴

Of course, tables can be sorted by a single indicator. Following the grouping approach, institutions are sorted alphabetically within groups – the ranking does not produce a league table, only groups. In the following example the institutions are sorted by the indicator 'research publication output'.

Table 7-2: Default table with three indicators per dimension; sorted by indicator 'research publication output

	Teaching & Learning			Research			Knowledge transfer			international orientation			Regional engagement		
	student staff ratio	graduation rate	qualification of academic staff	research publication output	external research income	citation index	% income third party funding	CPD courses offered	startup firms	international academic staff	% international students	joint international publ.	graduates working in the region	student internships in local enterprise	regional co-publication
Institution 2	🔴	🟡	🟡	🟢	🟢	🟢	-	🔴	🔴	🟡	🔴	🟡	🔴	🔴	🔴
Institution 4	🟢	🟡	🟡	🟢	🔴	🟡	🔴	🟡	🟡	🟢	🟡	🔴	🟢	🔴	🟡
Institution 1	🟡	-	-	🟡	🟢	🟢	🟢	🟡	🟢	🟡	🟡	🟢	🟢	🟡	🟢
Institution 3	🟡	🟢	🟢	🟡	🟢	🟡	🟡	🟢	🟡	🟡	🟢	🟡	🟡	-	🟡
Institution 7	🟡	🟡	🔴	🟡	-	🔴	🟢	-	🟢	-	🟡	🟡	🟡	🔴	🟡
Institution 8	🟢	🟡	🟡	🟡	-	🟡	🟡	🔴	🟡	🔴	🟡	-	🟡	🟢	🟢
Institution 9	🟡	🔴	🔴	🟡	🔴	🔴	🔴	🟡	🔴	🟢	🟡	🔴	🔴	🟡	🔴
Institution 5	🟡	🟢	🟢	🔴	🟢	🟢	🟡	🟢	🔴	🟡	🔴	🟢	🟡	🟢	-
Institution 6	🔴	🔴	🟡	🔴	-	🔴	🟢	🟡	🟢	🔴	🟢	🟢	🟢	🟡	🟢

In chapter 1 we discussed the necessity of multidimensional and user-driven rankings for epistemological reasons. Empirical evidence from the feasibility study strongly supports this view. The overview table above shows several institutions from the pilot sample and demonstrates that no institution performs in the top group (or bottom group) on all dimensions and indicators. While some institutions demonstrate average performance in many indicators others show a clear performance profile with marked strengths and weaknesses.

Users may examine one or more dimensions in depth, drilling down to the second layer of the table by clicking on a single dimension , e.g. 'Research', which will then display the complete list of all indicators in that dimension.

Table 7-3: Default table for one dimension

	Research				
	external research income	research publication	doctorate productivity	field normalised citation rate	highly cited research
Institution 8					
Institution 1		-	-		
Institution 3					
Institution 7					
Institution 9					
Institution 5					
Institution 2					
Institution 6					

7.3.2 Personalized ranking tables

The development of an interactive user-driven approach is a central feature of U-Multirank. Users have different views on the relevance of indicators included in a ranking and the tool recognizes this by allowing users to select the individual indicators they feel are relevant. This option is available both for the focused institutional rankings and the field-based rankings.

Personalized ranking implies a two-step process:

- First, users select a limited number of indicators, from one or more dimensions
- In a second step, users can specify the result table by choosing rank groups for each indicator selected (e.g. top level only; at least mid-table, all groups etc).

The following figure shows how users can select indicators.

Figure 7-2: User selection of indicators for personalized ranking tables

The 'green' column refers to top group only; the 'green and yellow' column refers to at least the middle group and the final column to all groups

The result will be a personalized ranking according to the selection of indicators by the user.

Table 7-4: Personalized ranking table

	Personalised ranking				
	international academic staff	research publication	doctorate productivity	student internships in	CPD courses offered
Institution 8					
Institution 1		-	-		
Institution 3					
Institution 9					
Institution 7					
Institution 5					
Institution 2					
Institution 6					

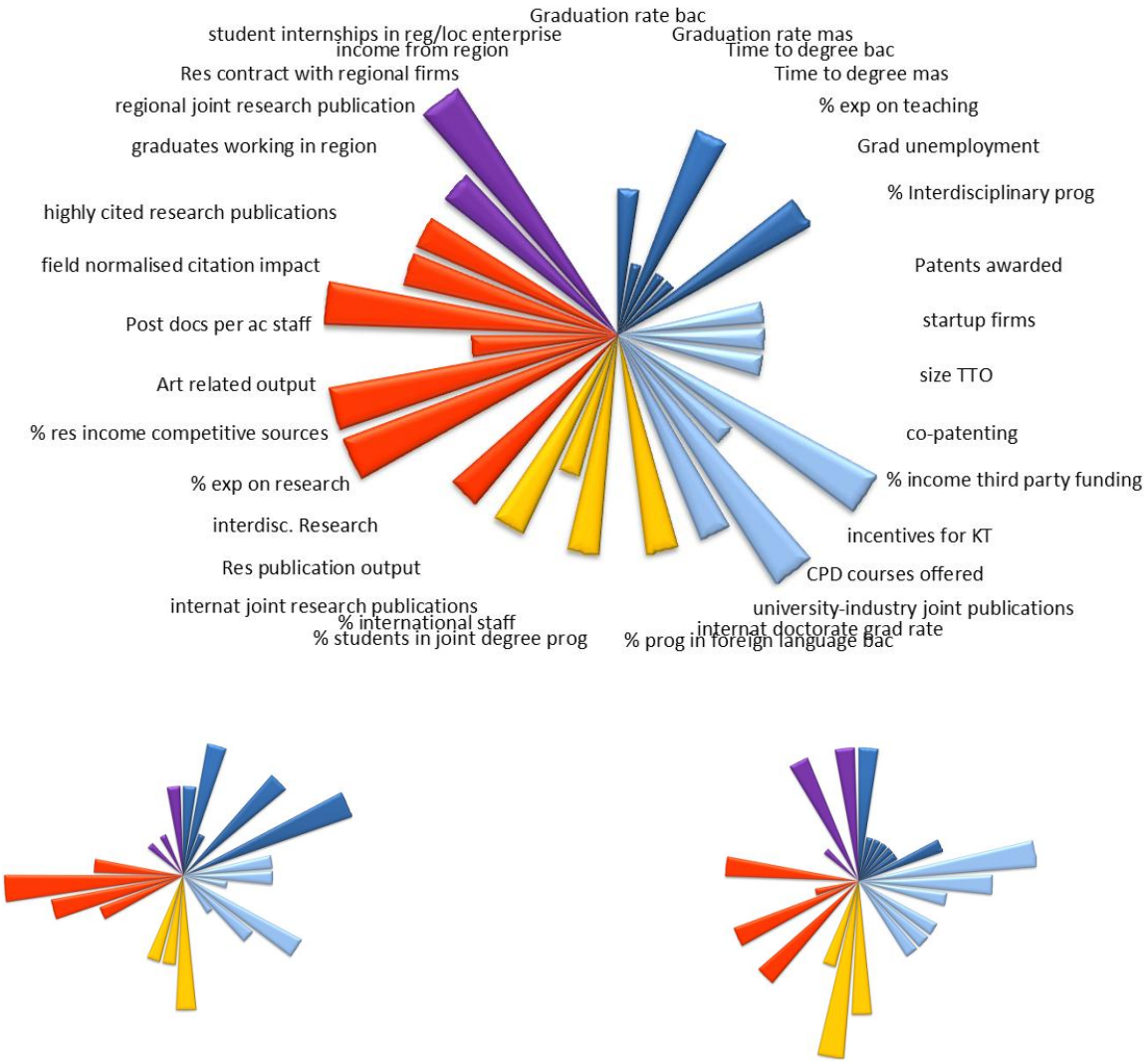
7.3.3 *Institutional results at a glance: sunburst charts*

Not all users will want to read a lengthy table when applying U-Multirank. An intuitive, appealing visual presentation of the main results will introduce users to the performance ranking of higher education institutions. Results at a glance presented in this way may encourage users to drill down to more detailed information.

Graphic presentations may help to convey insights into the institutional results 'at a glance' with the performance of the institution as a whole presented without being aggregated into one composite indicator.

The number of presentation modes should be limited, so that there is a recognizable U-Multirank presentation style and users are not confused by multiple visual styles. Four 'at a glance' presentation options were displayed and discussed at a U-Multirank stakeholder workshop and there was a clear preference for the 'sunburst' chart similar to the one used in U-Map. The colours symbolize the five U-Multirank dimensions, with the rays representing the individual indicators. In this chart the grouped performance scores of institutions on each indicator are represented by the length of the corresponding rays: the larger the ray, the better the institution performs on that indicator. As shown in Figure 7-3 different sunburst charts show different institutional performance profiles.

Figure 7-3: Institutional sunburst charts

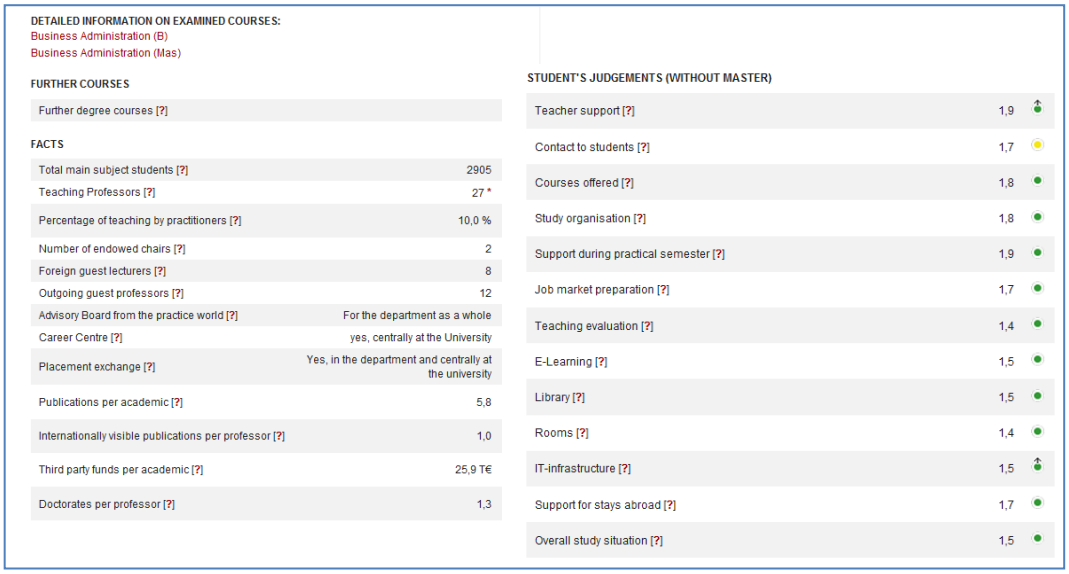


7.3.4 *Presenting detailed results*

In addition to the graphic presentation of the results of an institution, detailed information may also be presented in text formats.

An example is a detailed view on the results of a department (the following screenshot shows a sample business administration study program at bachelor and masters level). Here the user finds all indicators available for this institution – compared to the complete sample (the groups) – as well as additional descriptive contextual information (e.g. on the size of the institution/department). This kind of presentation can be made available on the three levels of institution, faculty/department (field) and program.

Figure 7-4: Text format presentation of detailed results (example)



DETAILED INFORMATION ON EXAMINED COURSES:
Business Administration (B)
Business Administration (Mas)

FURTHER COURSES

Further degree courses [?]	

FACTS

Total main subject students [?]	2905
Teaching Professors [?]	27 *
Percentage of teaching by practitioners [?]	10,0 %
Number of endowed chairs [?]	2
Foreign guest lecturers [?]	8
Outgoing guest professors [?]	12
Advisory Board from the practice world [?]	For the department as a whole
Career Centre [?]	yes, centrally at the University
Placement exchange [?]	Yes, in the department and centrally at the university
Publications per academic [?]	5,8
Internationally visible publications per professor [?]	1,0
Third party funds per academic [?]	25,9 T€
Doctorates per professor [?]	1,3

STUDENT'S JUDGEMENTS (WITHOUT MASTER)

Teacher support [?]	1,9
Contact to students [?]	1,7
Courses offered [?]	1,8
Study organisation [?]	1,8
Support during practical semester [?]	1,9
Job market preparation [?]	1,7
Teaching evaluation [?]	1,4
E-Learning [?]	1,5
Library [?]	1,5
Rooms [?]	1,4
IT-infrastructure [?]	1,5
Support for stays abroad [?]	1,7
Overall study situation [?]	1,5

7.4 Contextuality

Rankings do not and cannot provide causal analyses of their results. They are comparisons of performance results and offer information without claiming to be able to explain the differences in performance. Nevertheless, rankings have to take into account that contextual factors are highly relevant when comparing results (Yarbrough 2011; see also chapter 1). In general two types of context factors can be distinguished:

- Context variables affecting the performance of higher education institutions.
- Context factors that may affect decision-making processes of users of rankings (e.g. students, researchers) although not linked to the performance of institutions.

For individual users rankings reveal that there are differences in reality. For instance: for prospective students intending to choose a university or a study program, low student satisfaction scores regarding the support by teaching staff in a specific university or program is relevant information, although the indicator itself cannot explain the reasons behind this judgment.

Rankings also have to be sensitive to context variables that may lead to methodological biases. An example which has been discussed intensively (cf. Van Raan 2007) is the use of the publication of journal articles and article-based citations in institutional rankings.

Analytically, relevant context variables can be identified at different levels:

- The institution: context here can refer to the age, size and field structure of the institution.

- The (national) higher education system as a general context for institutions: this includes legal regulations (e.g. concerning access) as well as the existence of legal/official 'classifications' of institutions (e.g. in binary systems, the distinction between universities and other forms of non-university higher education institutions).
- The structure of national higher education and research: the organization of research in different higher education systems is an example. While in most countries research is largely integrated in universities, in some countries like France or Germany non-university research institutions undertake a major part of the national research effort.

A particular issue with regard to the context of higher education refers to the definition of the unit of analysis. The vast majority of rankings in higher education are comparing higher education *institutions*. A few rankings explicitly compare higher education systems, either based on genuine data on higher education systems, e.g. the University Systems Ranking published by the Lisbon Council[31], or by simply aggregating institutional data to the system level (e.g. the QS National System Strength Ranking). In this latter case global institutional rankings are more or less implicitly used to produce rankings of national higher education systems, thereby creating various contextual problems. Both the Shanghai ranking and the QS rankings for instance are including universities only. The fact that they do not include non-university research institutions, which are particularly important in some countries (e.g. in France, Germany), produces a bias when their results are interpreted as a comparative assessment of the performance or quality of national higher education and research systems.

U-Multirank addresses the issues of contextuality by applying the design principle of comparability (see chapter 2). In U-Multirank rankings are only created among institutions that have sufficiently similar institutional profiles. Combining U-Map and U-Multirank produces an approach in which comparable institutions are identified before they are compared in one or more rankings. By identifying comparable institutions, the impact of contextual factors may be assumed to be reduced.

In addition, U-Multirank intends to offer relevant contextual information on institutions and fields. Contextual information does not allow for causal analyses but it offers users the opportunity to create informed judgments of the importance of specific contexts while assessing performances. During the further development of U-Multirank the production of contextual information will be an important topic.

[31] See www.lisboncouncil.net

7.5 User-friendliness

U-Multirank is conceived as a user-driven and stakeholder-oriented instrument. The development of the concept, the definition of the indicators, processes of data collection and discussion on modes of presentation have been based on intensive stakeholder consultation. But in the end a user-driven approach largely depends on the ways the results are presented. In U-Multirank a number of features are included to increase the user-friendliness.

In the same way as there is no one-size-fits-all-approach to rankings in terms of indicators, there is no one-size-fits-all approach to the presentation of the results. The presentation modes should allow addressing different groups of users differently. According to the Berlin Principles, rankings should 'provide consumers with a clear understanding of all of the factors used to develop a ranking, and offer them a choice in how rankings are displayed.' (International Ranking Expert Group 2006; principle 15). U-Multirank, as any ranking, will have to find a balance between the need to reduce the complexity of information on the one hand and, at the same time, to offer detailed information that meet the requirements of specific users on the other.

U-Multirank wants to offer a tailor-made approach to presenting results, serving the information needs of different groups of users and taking into account their level of knowledge about higher education and higher education institutions. Basic access is provided by the various modes of presentation described above (overview tables, personalised rankings and institutional profiles). In addition access to and navigation through the web tool will be made highly user-driven by specific 'entrances' for different groups of users (e.g. students, researchers/academic staff, institutional administrators, employers) offering specific information regarding the results. Such a tailor-made approach implies different kinds and degrees of 'guidance' of users through the ranking processes.

In accordance with EU policies on eAccessiblity[32] barriers to access to the U-Multirank results and data will be removed as much as possible. This refers to a number of issues:

- Language: most 'expert users', i.e. users from within higher education will be able to use an English version of U-Multirank. In particular for 'lay users' (e.g. prospective students) the existence of various language versions of U-Multirank would increase usability. However, translation of the web tool and the underlying data is a substantial cost factor. But at least an explanation of how to use U-Multirank and the glossary and definition of indicators and key concepts should be available in as many European languages as possible.

[32] See http://europa.eu/legislation_summaries/information_society/l24226h_en.htm (retrieved on 10 May 2011).

- Visual access: in order to guarantee broad accessibility visual barriers relating to the color system of rank groups can be removed by introducing clear symbols at the same time.
- Free access: access to at least basic results should on the first hand be free of charge. A system that is – partly – subsidized by public funding should ideally be free for its users. In the long run the need to establish a self-sustaining system could imply the need to charge user fees for more detailed and advanced analyses. There will be a continuous trade-off between open access to results and a feasible business model to finance U-Multirank (see chapter 8).

Another important aspect of user-friendliness is the transparency about the methodology used in rankings. For U-Multirank this includes a description of the basic methodological elements (institutional and field-based rankings, grouping approach), a description of underlying data sources (e.g. self-reported institutional data, surveys, bibliometric data, patent data) and a clear definition and explanation of indicators (including an explanation of their relevance and what they are measuring). In an online ranking this information should be available electronically, too. An electronic description of the methodology can be linked to the presentation of results (e.g. by using hyperlinks) and hence increase the understanding of the ranking substantially.

In the end the user-friendliness of a ranking tool cannot be assessed *a priori*. Tracking ranking use is important. How do users choose to navigate through the web tool? What indicators are selected most frequently in personalized rankings? How deeply do users examine the results and where do they stop navigation? Tracking of user behaviour can be systematically built into the implementation of the web tool and by doing so can help to increasingly adapt the tool to the needs of users.

8 Implementing U-Multirank: the future

8.1 Introduction

An important aspect in terms of the feasibility of U-Multirank is the question of implementing the system on a widespread and regular basis. In order to address this question we need to look for appropriate organizational models, conceptualize functioning organizational units, ensure institutional sustainability and find funding sources to cover the costs the system. This chapter deals with all these issues of implementation and institutionalization for which we will develop several scenarios and options. It is clear that the implementation of U-Multirank is a dynamic and only partially predictable process which will not reach a stable outcome for a number of years. Nevertheless, the stages of development have to be planned and if necessary readjusted and we must differentiate between a two-year pilot phase and a longer-term implementation/institutionalisation of U-Multirank.

One of our basic suggestions regarding transparency in higher education and research is the integration of U-Map and U-Multirank. Both are complementary transparency instruments and they should be implemented in a consistent manner. Therefore, many of the conclusions regarding the operational implementation in the final U-Map report (see www.u-map.eu) are also valid for U-Multirank. The link between the two projects has been created by guaranteeing the use of U-Map data for the selection of comparable (and therefore 'rankable') institutions.

8.2 Scope: global or European

The pilot test showed some problems with the inclusion into U-Multirank of institutions from specific countries outside Europe. Clearly, with participation in U-Multirank on a voluntary basis higher education institutions will have to be convinced of the benefits of participation. This leads to the question of the scale of international scope that U-Multirank could and should attain.

We would argue that U-Multirank should aim to achieve a relatively wide coverage of European higher education institutions as quickly as possible during the next project phase. In Europe the feasibility and willingness to participate proved to be high and in Europe a full complement of institutional profiles is preferable in order to be able to address the diversity of European higher education. But U-Multirank should remain a global tool. There are institutions all over the world interested in benchmarking with European universities; the markets and peer institutions for European universities are increasingly becoming global; and the impression that the U-Multirank instrument is

only there to serve European interest should be avoided. The pilot study proves that U-Multirank can be applied globally. Based on the pilot results we suggest that the extension beyond Europe could best be organized systematically and should not represent just a random sample. From outside Europe, the necessary institutions should be recruited to guarantee a sufficient sample of comparable institutions of different profiles. For instance it could be an option to try and integrate the research-oriented, international universities scoring high in traditional rankings.

When this strategy leads to a substantial database within the next two years, recruitment could be reinforced, at which point the inclusion of these important peer institutions will hopefully motivate more institutions to join U-Multirank.

Another aspect of scope is the timing of rankings: how often should rankings take place? The frequency of data collection is always a compromise between obtaining the most up to date information and the workload that data-gathering imposes on the institutions. For the institutional ranking data collection would probably take place via a full update for instance every two or three years. We suggest a rolling system for the field-based ranking.

There is no definitive answer to the question of how many fields there are in international higher education. ISCED (1997) includes nine broad groups, such as humanities and arts, science, and agriculture. Within these groups, there are 25 fields of study.[33] UNESCO is considering further dividing these fields by adding sub-categories in order to generate more precise statistical information.

Based on our pilot project we believe that it is feasible to add five new fields in each of the first three years of continued implementation of U-Multirank. If the rankings were updated on a three-year rolling schedule this would allow coverage of 15 fields. At that stage a better informed decision about the feasibility of extending the coverage of the rankings to further fields could be taken. In terms of selecting which fields to add it would make sense to focus on fields with significant numbers of students enrolled and to ensure that the initial 15 fields span all nine broad subject groups.

8.3 Personalized and authoritative rankings

As explained before, the 'heart' of U-Multirank is the idea of creating a user-driven, flexible tool to obtain subjective ranking that are relevant from the perspective of the individual user. This implies that data updates would not lead to the publication of a

[33] The Frascati manual has a similar structure but there are significant differences between its broad groups and fields of study and those of ISCED. This is an area where harmonisation would yield many benefits.

final, static ranking but would only feed into the database, allowing the user to rank on the basis of the most current information. However, with U-Multirank it is also possible to create so-called 'authoritative' ranking lists from the database. A ranking producer aiming to create a specific ranking for dissemination to a wider audience may choose a meaningful subset of institutions and publish an 'authoritative' ranking, adding their own authority and clear explanations to the ranking list. There are various options available:

- An authoritative ranking could be produced by a specific association of higher education institutions. For instance international associations or consortia of universities (such as CESEAR, LERU or ELIA) might be interested in benchmarking or ranking their 'members'.
- An authoritative ranking could be produced from the perspective of a specific stakeholder or client organization. For instance, an international public organization might be interested in using the database to promote a ranking of the international, research-intensive universities in order to compare a sample of comparable universities worldwide.
- A national or international private media company might be interested in producing a ranking on one or more dimensions, choosing the indicators relevant to a broader audience (or their specific target audience).

These and other examples imply that an organization (or group of organizations) offers its own specific selection of dimensions and indicators to provide some perspective for a wider audience. As we have argued throughout this report, epistemologically speaking there is no argument for claiming that one set of dimensions and indicators is better or more relevant than others. In this context, however, these specific organizations indeed claim a certain 'authority' for their selection.

The question could be raised whether the option of producing 'authoritative' rankings should be followed during the next project phase. On the one hand, this might be an important means of generating revenue from database-derived products. On the other hand, in the first phase of implementation, U-Multirank should be perceived by all potential users as relevant for their individual needs. We recommend exploring the option of producing authoritative rankings in the next project phase by creating authoritative rankings on a pilot basis with one or two international associations of higher education institutions and by conceptualizing one or more authoritative rankings with interested public and/or private partners. The major focus of the coming implementation phase, however, is on establishing the flexible web tool.

8.4 The need for international data systems

U-Multirank is not an isolated system, but has relevant links (not only to U-Map but also) to other instruments and developments in the following two respects.

The quality of the results of any transparency tool depends to a large extent on the availability of relevant and appropriate data. Our pilot study has shown that there are still challenges in this respect.

There is a strong need for a European data system, with institution and field data, preferably with clear relationships to other data systems in the world (such as IPEDS). The EC should promote the development and harmonization of these institution and field-based indicators and their definitions and calculation methods in national statistics, and should continue the data collection in a follow-up of the recently finalized EUMIDA project. The development of the European database resulting from EUMIDA should take into account the basic data needs of U-Multirank. This would allow the pre-filling of institutional questionnaires with available data and would substantially reduce the workload for the institutions.

Some specific recommendations regarding the further development of the EUMIDA database can be made:

- First, there are some elements which seem to be easily collectable by EUMIDA, such as staff data (the proper and unified definition of full-time equivalents and the specification of staff categories such as 'professor' is an important issue for the comparability of data), or data related to students and graduates. EUMIDA could contribute to improve the data situation regarding employment-oriented outcome indicators. An open question is how far EUMIDA is able to go into field-specific data; for the moment pre-filling from this source seems to be more realistic for the institution-level data than for field-based data.

- A second aspect of integrated international data systems is the link between U-Multirank and national ranking systems. U-Multirank implies a need for an international database of ranking data consisting of indicators which could be used as a flexible online tool in order to create personalized rankings by users (according to the user's preferences). This database is a crucial starting point to identify and rank comparable universities. Developing a European data system and connecting it to similar systems worldwide will strongly increase the potential for multidimensional global mapping and ranking.
  Despite this clear need for cross-national/European/global data there will be a continued demand for information about national/regional higher education systems, in particular with regard to undergraduate higher education. Although

mobility of students is increasing, the majority of – in particular undergraduate – students will continue to start higher education in their home country. Hence field-based national rankings and cross-national regional rankings (such as the CHE ranking of German, Dutch, Austrian and Swiss institutions) will keep their relevance. A 'bottom-up' system could work with countries or groups of countries in order to develop national field-based rankings based on common standards. Such a system could provide a common platform for the presentation of those different national/regional rankings. Furthermore it could also be used as a base for an international database and international rankings; more and more countries could come in step by step with a core set of indicators used in each ranking. The national rankings could refer to specific national higher education systems and at the same time provide a core set of joint indicators that can be used for European and global rankings.

We currently observe that there is also interest from nation-states (as e.g. the Netherlands, Flanders, Portugal and Estonia) or groups of countries (e.g. the Nordic countries) in joining the U-Map European mapping exercise. In Spain we have the example of Fundacion CYD planning to implement a field-based ranking system based on U-Multirank standards. The CHE ranking sponsored by the Bertelsmann Foundation already covers Germany, Austria, Switzerland and the Netherlands. These initiatives reflect a 'bottom-up' approach to building cross-national data systems. They can (and hopefully will) be expanded thus creating an increasing set of data systems to be combined into a joint database.

How to deal with the top-down and bottom up-approach? The only reasonable solution is to combine the two as complementary approaches and to use national institutional and field-based rankings for pre-filling. In its operational phase the U-Multirank unit should develop standards and a set of basic indicators that national initiatives would have to fulfil in order to become part of the international cooperative structure. In the 'bottom-up' approach national rankings could feed their data into the international database, the U-Multirank unit will be able to pre-fill the data collection instruments and has to fill the gaps to attain European or worldwide coverage. At the same time activities based on the top-down approach might help to make the system known and to develop trust and credibility. Top-down rankings would also become less expensive to implement if they could use existing national data and data collection infrastructures. Also, gaining sponsorship for the system could sometimes be easier starting from the national level; initial discussions with foundations revealed a greater willingness to engage financially in ranking activities at the national level. Finally, the existence of bottom-up elements would enable different organizations providing a ranking to join a larger 'ranking family' and would avoid monopoly solutions.

8.5 Content and organization of the next project phase

Most of the previous discussion referred to the medium- and long-term implementation. But there are also some clear next steps in the implementation process which should be taken relatively quickly. These steps could be taken within two years in a project structure similar to that of the feasibility study, again supported by European Commission funding. The work packages for the next phase of implementation should be:

Finalisation of the various U-Multirank instruments

1. *Full development of the database and web tool*. The feasibility study showed how the flexible, user-driven ranking could work; now the real system has to be created, populated with data and tested, and has to start running. The prototypes of the instrument will demonstrate the outcomes and benefits of U-Multirank.
2. *Setting of standards and norms and further development of underdeveloped dimensions and indicators.* A core set of indicators should be defined, definitions of data concepts should be fixed, standardized elements of data collection tools should be developed. In the feasibility study we found indicators and dimensions where the data collection was difficult, but they have high relevance and we discovered sufficient potential to develop adequate concepts and data collection methods. These parts of the ranking model should be further developed.
3. *Update of data collection tools/questionnaires according to the revision and further development of indicators and the experiences from the U-Multirank project.* Depending on the further development of indicators and their operationalization, the data collection instruments have to be adapted. A major issue is to design the questionnaires in a way that reduces administrative burden for the institutions as far as possible.

Development of pre-filling in EU+ countries

4. *Further development of pre-filling.* In the first round of U-Multirank pre-filling proved difficult. The testing of national data systems for their pre-filling potential and the development of suggestions for the promotion of pre-filling are important steps to lower the costs of the system for the institutions. A link to the development of a European higher education data system (EUMIDA) should be explored; coordination of all relevant EC projects should be part of the next phase. In addition, the test of a coordinated approach between one or several national field-based rankings and the international U-Multirank database should be realized.

Roll-out of U-Multirank across EU+ countries

5. *Invitation of EU+ higher education institutions and data collection.* Within the next two years all identifiable European higher education institutions should be invited to

participate in the institutional as well as in the three selected field-based rankings. The objective would be to achieve full coverage of institutional profiles and have a sufficient number of comparable institutions. If we take into account the response rate of institutions in the pilot phase the inclusion of 700 institutions in the institutional and 500 in each field-based ranking appears realistic.

6. *Targeted recruitment of higher education institutions outside Europe.* In order to guarantee a global scope to the project, institutions with relevant profiles should be approached.
7. *Conceptualization of specific authoritative rankings.* The combined U-Map/U-Multirank approach should be tested further by developing the means to produce the first authoritative ranking lists for universities with selected profiles. Rankings could be conceptualized for two different profiles, for instance the profile of research orientation and a high degree of internationalization (international research intensive universities) and the profile of a strong focus on teaching and learning and a clear regional orientation. Furthermore, partners should be identified from international groups and alliances of higher education institutions willing to establish internal benchmarking processes and publish rankings of their membership.

Business Model

8. *Business plan and marketing.* If the objective is to establish U-Multirank as largely self-sustainable, a business plan is required. It could be a good idea to involve organizations with professional business expertise in the next project phase in order to work out a business plan, and to analyze the revenue-generating potential, development of marketable products, pricing issues etc. The business plan must address a fundamental contradiction: the user-driven approach imbues U-Multirank with strong democratic characteristics and a role far from commercial interests, while at the same time a certain degree of commercialization becomes inevitable if complete funding from non-profit sources is unrealistic.
9. *Formal institutionalization of the U-Multirank unit.* During the next project phase an operational organization to implement U-Multirank will need to be created and a governance and funding structure established, the advisory bodies will need to start working and the legal structure for the operational unit must be specified.

Communication

10. *Communication and recruitment drive.* The features of and opportunities offered by U-Multirank need to be continuously communicated. Since the success of U-Multirank requires institutions' voluntary participation a comprehensive promotion and recruitment strategy will be needed, requiring the involvement of many key players

(governments, European Commission, higher education associations, employer organizations, student organizations).

11. *User-friendliness of the instrument.* A crucial issue related to communication is the user-friendliness of U-Multirank. This could be guaranteed by the smoothness of data collection and the services delivered to participants in the ranking process. But user-friendliness also deals with the design of the web tool, taking into account the differing information needs and knowledge about higher education of specific user groups (for instance secondary school leavers versus higher education decision-makers). A user-friendly tool needs various levels of information provision, understandable language, clarity of symbols and explanations, assisted navigation through the web tool and feedback loops providing information about users' preferences.

Bringing the 11 work packages and the resulting products together in a feasible schedule leads to the following project structure (assuming the next project phase starts 01/2012 and ends 12/2013):

Table 8-1: Elements of a new project phase

Work package	Products	Deadline
Database and web tool	• Functioning database • Functioning web tool prototype	06/2012
Standards and norms	• Description of standards and norms • Final data model	06/2012
Finalized collection tools	• Collection tools	06/2012
Pre-filling	• Planning paper on pre-filling opportunities (including EUMIDA cooperation) • Pre-filled questionnaires • Coordination with national rankings	06/2012 12/2012 12/2012
Roll-out	• Invitation • Targeted non-European recruitment • Data collection • Data analysis and publication	06/2012 09/2012 03/2013 06/2013
Specific focused rankings	• Two rankings conceptualized • One benchmarking exercise	12/2013 12/2012

	with alliance	
Formal institutionalization, business plan, user-friendliness and communication	• Advisory boards work • Consortium, formal organization and business plan (including funding structure) established • Development of marketable products	06/2012 12/2013 12/2013

The 11 elements form the potential content of the next U-Multirank project phase, transforming U-Multirank from a feasible concept to a fully developed instrument already rolled out and ready for continuous operation.

8.6 Criteria and models of implementation

An assessment of the various options for the organizational implementation of U-Multirank requires a set of analytical criteria. The following criteria represent notions of good practice for this type of an implementation process such as governance, organization, funding structures etc. The criteria were derived from the analytical findings of the feasibility study, from the stakeholder consultation process, from the earlier U-Map study (see www.u-map.eu) and from the standards set out in the Berlin Principles:

Credibility: The transparency tool must have the trust of participating institutions and other stakeholders. This means that the organization managing the instruments must be accountable and subject to continuous evaluation and assessment. Credibility will ensure legitimacy and acceptance of the instruments. The organizational structure also contributes to credibility, for instance if checks and balances are assured, through oversight by supervisory bodies etc.

Inclusiveness: The ranking must be open to recognized higher education institutions of all types and from all participating countries, irrespective of their membership in associations, networks or conferences.

International orientation: Implementation of the transparency tool must involve actors and expertise from different countries and international and supranational support structures. In order to be internationally credible, the implementation should not be the prerogative of an institution with national character, and the structure should take into consideration, in a credible manner, the various linguistic, cultural, economic and historical contexts in different countries.

Professional approach: The ranking must be run by a professional organization with expertise in large-scale data analysis and in transparency tools. This will guarantee high standards of the planning, implementation, communication and further development of the instruments, hence contributing to the transparency tool's impeccable reputation, which is essential to its success. A professional approach will guarantee the ongoing methodological development of ranking tools.

Sustainability: The implementation must be properly funded, carried out on the basis of a secure long-term commitment in order to ensure the capacity needed to carry out the work at the required high level. Funding could include a mix of different sources, including non-governmental sources among others.

Efficiency: A multidimensional ranking implies high workloads for institutions and ranking providers. Therefore, efficiency in data collection is important. This criterion also refers to an efficient link between national rankings and the international ranking tool. In addition, efficiency refers to the coordination of different European initiatives to create international databases (such as E3M, EUMIDA).

Independence: The ranking tool must be administered independent of the interests of higher education institutions or representative organizations in the higher education and research sector. The ownership and organizational structure must guarantee that the ranking results are not influenced by political or strategic interests. The implementation has to separate ranking from higher education policy issues such as higher education funding or accreditation. Independence is also enhanced when the organizational structures do not lead to data monopolies.

Service orientation: A key element of U-Multirank is the flexible, stakeholder-oriented, user-driven approach. The implementation has to ensure this approach, for instance by integrating stakeholders into consultation structures, creating information products for stakeholder needs and service-oriented communication processes.

In general, the involvement of relevant actors in both the implementation of U-Multirank and its governance structure is a crucial success factor. In order to underpin the application of this ranking tool with sufficient credibility and trustworthiness, those parties taking responsibility for the governance of U-Multirank should be broadly accepted by stakeholders. Those who will be involved in the implementation should allow their names to be affiliated with the new instrument, promote its development, supply a part of the funding and take responsibility in the governing bodies of the organizational structure.

We identified four basic options for responsibility structures for U-Multirank:

Commercial model: In this model (a consortium of) private, for-profit organizations would run the instrument with for-profit objectives. Products and services would be made available to users at market-based tariffs. The strategy, use of the instrument and its further development would be driven by market demands. Potential organizations could be newly founded, but existing institutions could also take on the role, e.g. media companies (interested in publishing rankings), consulting companies in the higher education context and data providers (such as the producers of bibliometric databases).

Government model: In this model, governments would use their authority over higher education to organize the rankings of higher education institutions. As the tool to be developed has a European or worldwide character, it would be owned either at the supranational level by the European Commission or within the framework of an inter-governmental agreement.

Stakeholder model: In this model, major stakeholders, i.e. student organizations and associations of higher education institutions, would be responsible for the operation of the transparency instrument.

Independent non-profit model: In this model, an existing or new cross-national/international organization (or alliance of organizations) independent of government or direct stakeholder interests would act as principal of the transparency tools. The organization would work under non-profit conditions and would have to find a funding structure covering the cost. Independent organizations could be non-governmental (research) institutions (such as the CHERPA network) or other NGOs/think tanks. One possibility, probably with the highest degree of independence, could be a consortium of foundations from different countries.

The following tables present the pros and cons of the different model options, taking into account the criteria.

Table 8-2: Pros and cons of four alternative implementation models

COMMERCIAL MODEL	
PRO	CON
• Service and customer orientation has to be high, adaptation to user needs required. • Independence from direct political influence. • Profit orientation is a good incentive to be efficient.	• Only services with sufficient demand are offered (market dependence), multidimensional approach and inclusiveness endangered. • Continuation of system depends on market conditions and their volatility, danger for sustainability. • Profit orientation endangers quality and credibility. • Doubts about financial feasibility, because if HEI experience high workloads with data collection they expect free products in return and are not willing to pay for basic data analysis. • Doubts about commitment to social values of European higher education area (e.g. no free access for student users?). • International character is not guaranteed. • If the system has to be profitable, the cost of surveys will become a major issue and more expensive dimensions of the ranking might be neglected (reducing the added value to

GOVERNMENT MODEL	
PRO	CON
• Government authority ensures participation and inclusiveness. • International approach is possible. • High credibility of the actors and high sustainability if there is a long-term decision to establish an operational unit.	• Independence from political interests could not be ensured. • Difficult to keep system separate from steering mechanisms such as funding etc. • Results of rankings will create direct pressure on national governments and enhance the tendency to influence them. • Service orientation might not be the primary interest in a state-run system.

STAKEHOLDER MODEL	
PRO	CON
• High legitimacy and acceptance among stakeholders (included). • Good chance for international orientation.	• If not all stakeholders are represented inclusiveness becomes difficult. • Inefficiency because of difficulties to find a common ground between stakeholders. • No independence from stakeholder organization interests could be ensured, therefore problems with credibility from the point of view of the end user.

INDEPENDENT NON-PROFIT MODEL	
PRO	CON
• Institutions with strong funding base such as foundations enhance sustainability. • Independence and non-profit orientation are excellent bases for trust into system and high credibility. • Non-profit organization can be linked with commitment to social values of European higher education area. • The idea of international alliances ensures international orientation. • Foundations have interest in inclusiveness and in efficient use of their resources.	• Institutions with weak funding base such as research institutes endanger sustainability.

In the figure below, we present a summary of our assessment of the four models against the criteria for operational implementation. Professionalism could be attained by establishing a professional operative unit regardless the realized ownership model, so it is not included in the comparative assessment.

Figure 8-1: Assessment of the four models for implementing U-Multirank

Criteria / Model	Inclusiveness	International orientation	Independence	Professionalism	Sustainability	Efficiency	Service orientation	Credibility
Commercial	-	-	+/-		-	+	+/-	-
Government	+	+	-		+/-	+/-	-	+/-
Stakeholder	-	+	-		+/-	+/-	+/-	+/-
Independent non-profit	+/-	+	+		+/-	+	+/-	+

The figure shows that there is not a clear 'best solution' among the four models. Therefore, our recommendation is not to opt for one of the models but to combine the different logical extremes to a 'mixed model' which includes elements of the different options.

8.7 Towards a mixed implementation model

We suggest that the starting point for a mixed model is the international, independent, non-profit model. It is independent both from higher education institutions (and their associations) and from higher education funding bodies/politics. It keeps ranking separate from steering and funding instruments in higher education. It can offer a non-commercial character to the instrument, and it can guarantee external supervision of the implementation and broad and open access to the results. As will be shown below it also allows the combination of different funding sources.

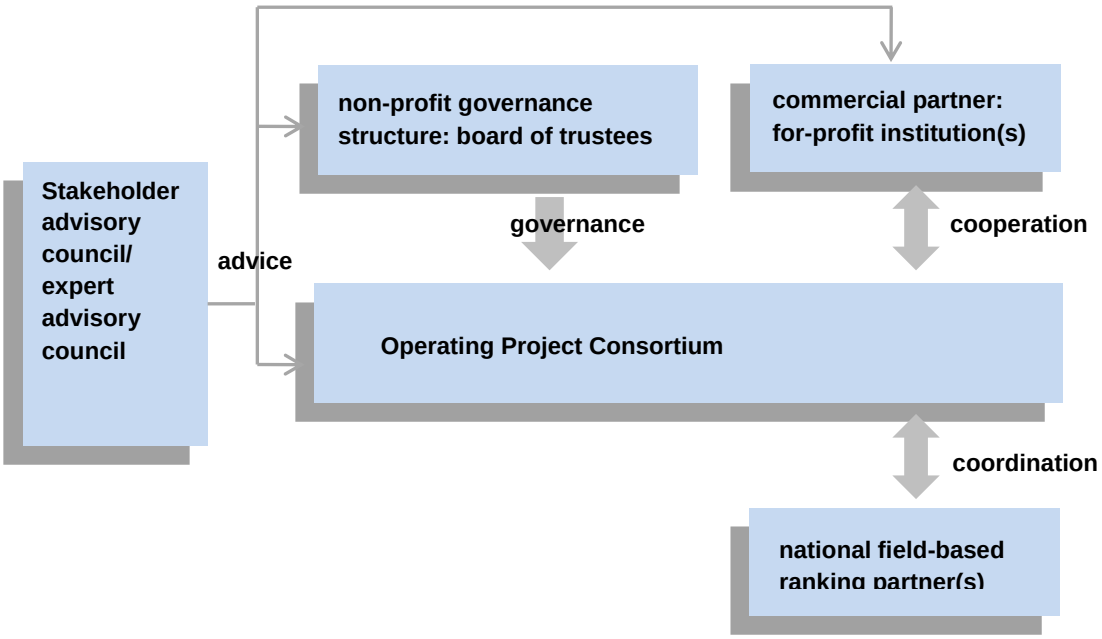
At the same time we see the models presented in section 8.6 as not incompatible alternatives. The best solution could be to use the independent model as the basis, but to combine it with elements of the other models. The stakeholder model should be recognized in so far as an advisory board could guarantee the connection to relevant groups of stakeholders.

As a realistic funding model seems to be impossible without commercial elements a partial inclusion of the commercial model is inevitable. The business plan has to include opportunities to charge fees, sell products and involve commercial partners. Products for universities could be created, but the pricing policy mustn't destroy the willingness to participate. A suggestion would be to organize the implementation of U-Multirank in such a way that basic ranking results can be provided for free to the participating institutions, but more sophisticated analyses would incur fees.

Government elements could be involved if governments decide to support and fund a ranking of their national systems; this seems to be an acceptable form of partial government ownership.

Following the general line of the mixed model, a specific organizational structure has to be found. Using a pragmatic approach the short-term model (for the next two years) has to differ from the long-term perspective. We believe that it is not reasonable in the initial phase of implementing U-Multirank to establish a new professional organization for running the system. Once the extent of participation of higher education institutions is known, this option could be considered. The assumption is therefore that rankings would be operated (initially) on a project basis by existing professional organizations with a strong involvement of both stakeholder and expert advisory bodies. There should be a next U-Multirank project phase before a ranking unit is established.

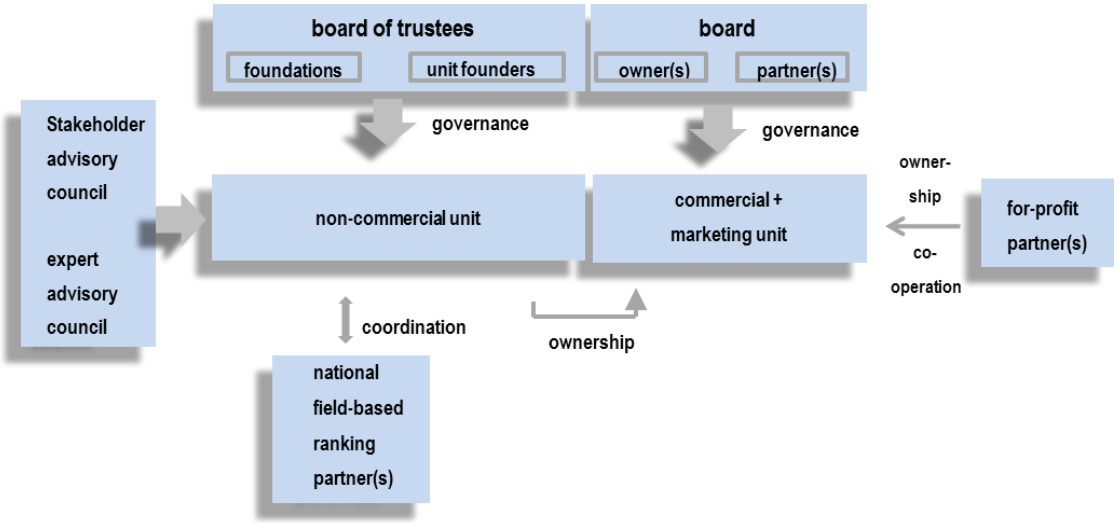
Figure 8-2: Organizational structure for phase 1 (short term)



We suggest that during the next two years of the project phase the current project structure of U-Multirank should be continued. The board of trustees will form a strategic, coordinating and decision-making body for the development of the transparency tool. The directors of the project partner entities should be represented in the board and a foundation or a similar relevant institution should be involved for future governance structures. Stakeholder and expert advisory councils should be installed in a form that could continue to operate after the two years' project phase. In order to support the development of a viable business plan a partnership with professional, commercial organizations experienced in business planning should be sought. The coordination with national field-based rankings should take place in pilot collaborations with national ranking initiatives.

After the first two-year period of cooperation, the future of the implementation of a formal organization will need to be decided. Then governance could be taken over by a more structured consortium of non-profit organizations/foundations and the operative unit could become a formal organization with legal status. The professional organizations responsible for the first phase could establish the ranking unit as a joint venture with the stakeholder and expert advisory structure remaining in place. This structure also allows the commercial unit to operate as a joint venture with for-profit partners. The operating units will function with clear tasks and loyalties.

Figure 8-3: Organizational structure for phase 2 (longer term)



The finalization of a long-term organizational structure should be one of the tasks for the implementation phase within the next two years.

8.8 Funding U-Multirank

The following analysis of cost factors and scenarios is based on the situation of running U-Multirank as an established system. Costs have been estimated based on this projection but will not become part of the final report. The cost estimations showed that U-Multirank is an ambitious project also in financial terms, but in general it seems to be financially feasible.

A general assumption based on the EU policy is that U-Multirank should become self-sustainable without long-term basic funding by the European Commission. EC contributions will decline over time and new funding sources will have to be found. However, from our calculations it became clear that there is no single financial source from which we could expect to cover the whole costs of U-Multirank; the only option is a diversified funding base with a mix of financial sources. If U-Multirank is not dependent on one major source a further advantage lies in the distribution of financial risks.

The cost factors are first of all related to the necessary activities involved in the production of ranking data:

- Methodological development and updates
- Communication activities
- Implementation of (technical) infrastructure
- Development of a database
- Provision of tools for data collection
- Data collection (again including communication)

- Data analysis (including self-collected data as well as analysis based on existing data sets as e.g. bibliometric analysis)
- Data publication (including development and maintenance of an interactive web tool)
- Basic information services for users
- Internal organization (advisory bodies, board).

It is difficult to calculate the exact running costs associated with U-Multirank because these depend on many variables. The major variable cost drivers of U-Multirank are:

- The number of countries and institutions involved. This determines the volume of data that has to be processed and the communication efforts.
- The number of countries/institutions which deliver data for free through a bottom-up system (this avoids costs).
- The number of fields involved. To limit cost a ranking could not cover all fields with sufficient size but could limit itself to the most studied fields.
- The surveys that are needed to cover all indicators outlined in the data models of U-Multirank. The cost depends not so much on a single indicator or dimension but more on the required surveys. Major cost factors are for instance the realisation of student and graduate surveys or the use of databases charged with license fees, e.g. bibliometric and patent data. An additional factor is the technological environment for the surveys, for instance a student survey is much more expensive if universities have no e-mail-addresses of their students, requiring students to be addressed by letters.
- The frequency of the updating of ranking data. A multidimensional ranking with data from the institutions will not be updated every year; the best timespan for rankings has to take into account the trade-off between obtaining up to date information and the workload for the institutions and the costs of updating data for the operative unit.

For the different steps we could identify the relevant cost factors, some fixed, some variable:

Table 8-3: Fixed and variable cost factors

STEP	FIXED COST FACTOR	VARIABLE COST FACTOR
Methodological developments and updates	• Staff demand	• Cycle of revision/update of concepts • Intensity of stakeholder involvement
Communication activities	• Staff demand	• Number of countries and institutions covered • Intensiveness of communication (written only, electronic, workshops etc)
Implementation of (technical) infrastructure	• Basic (technical) infrastructure, incl. IT	• Indicators/databases used (e.g. license costs)
Development of a database	• Staff • Basic IT costs	
Provision of tools for data collection	• Staff • Basic IT costs (incl. online survey systems)	• Comprehensiveness of set of indicators and databases
Data analysis	• Staff	• Number of countries and institutions covered • Range of indicators and databases • License fees of databases (e.g. bibliometric)
Publication	• Staff • Basic IT costs	• Features of web tool to present results
Information services for users	• Staff • Basic IT costs	• Number of countries and institutions covered • Range of indicators and databases • Scope of information services
Internal organization	• Costs for internal communication/ meetings	• Size of operative unit*

* The size of the operative unit itself is dependent on all factors listed above.

The major fixed cost elements in overview include:

- Basic ranking team: A basic team is needed to get the system running. The assumption is that we need at least four people – (1) a leading senior person with experiences in the fields of ranking and performance measurement and in stakeholder communication processes (as head of the project/unit); (2) two junior staff members with experiences in statistics, empirical research, large-scale data collection, IT; (3) secretarial support.
- Information technology support: further development and implementation of the on-line tool and related software development.
- Marketing and communication: the design and development of information packages on ranking and the dissemination of the outcomes as well as the staff time needed to do this.
- Research and analysis: the instruments will need to be carefully monitored, refined and analysed and the results researched and reported on. Staff time is the key cost.
- Meeting costs for the Board and the Councils: honoraria and travel and subsistence costs.

All other cost factors are variable and depend on the size and number of the ranking exercises.

After looking at the cost structures, the next step would be an analysis of the funding potentials and how costs might possibly be covered from different sources.

The intention of the European Commission is to develop U-Multirank into a self-sustaining instrument, requiring no EU funding after its implementation phase. Therefore, the objective here is to describe ways to attain self-sustainability. Nevertheless the European Commission should consider the option of continued support of part of U-Multirank's basic funding in the long run, which could serve to motivate other sponsors to co-fund the activities and ensure a formal role for the EC as a partner in U-Multirank. To promote transparency and performance in European higher education by establishing a transparency tool could be a long-term task of the EC. For instance the EC could take on the role of promoter of students' interests and could see the delivery of a web tool free of charge to students as its responsibility. To ensure students' free access to U-Multirank data the EC could provide – also in the long run – direct funding of user charges that would otherwise have to be imposed upon the students.

There are a number of opportunities to find funding sources for U-Multirank:

a) Basic funding by the governing institutions in form of a lump sum. This is realistic for the government model (e.g. basic funding by EU) and for the

independent, non-profit model (basic funding by interested national and international non-profit organizations that want to become the owners).

b) Funding/sponsorship from other national and international partners interested in the system.
c) Charges from the ranking users (students, employers etc.).
d) Project-based funding for special projects, for instance new methodological developments, rankings of a particular 'type' of institutions.
e) System and/or institutional 'subscription'.
f) Prices for derived products, such as special analyses of ranking data, to cross-subsidize the instruments.
g) Financial contributions from media partners publishing the results.
h) Non-financial contributions from third parties, such as free data provision.
i) Free provision of data from national mapping and ranking systems (bottom-up approach).

It is quite clear that funding from just one source is not realistic. Discussions with potential funders so far have shown that the funding of U-Multirank has to rely on a mix of income streams. But not all funding sources are available for all ownership models. The following table shows potential funding sources for the different ownership models.

Table 8-4: Funding sources

OWNERSHIP MODEL	POTENTIAL MAJOR FUNDING SOURCES	POTENTIAL ADDITIONAL FUNDING SOURCES
Commercial	Private funding, prices for products, contributions from media partners	(c), (d), (e), (f),
Government	Governmental funding	(c), (d), (f), (g) (h) (i)
Stakeholder	None	(b), (d), (e), (f), (g), (h), (i)
Independent non-profit	Basic funding by owners	(b), (d), (e), (f), (g), (h), (i)

Again, the advantages of the independent non-profit model (in a wider understanding including commercial elements) become apparent: together with the stakeholder model it has the broadest set of funding options, and in contrary to the stakeholder model also a clear potential basic funding source. If it is combined with the commercial model all relevant funding options are available.

The funding scenarios could be further specified and illustrated by examples. From the cost scenarios above we take the required funding volume and show these examples

illustrating plausible funding shares from different sources. The scenarios have to clarify who will pay the basic fixed costs of the ranking unit and who will pay the variable costs for ranking/data collection. The scenarios try to develop a medium-term perspective. We assume the realization of the independent, non-profit solution.

The first scenario is one where the principals of the ranking pay the major cost share.

Table 8-5: Funding scenario 1

COST FACTOR	COST SHARING
Basic fixed costs	100% principals
Rankings	50% principal 50% media partner, data providers, publishing companies

In this model, the non-profit principals of the system would contribute more than half of the funding, the rest would come from the market by involving partners who benefit in some way from the cooperation, such as media partners who publish rankings or data providers who benefit from being positioned in the ranking field. The principal's funding share could also include a contribution from the EC if there was a decision to engage in the long run.

The second scenario has a focus on user charges. We assume that a major part of the revenue comes from participation fees of higher education institutions, probably paid for them by national institutions (foundations, national governments, associations of institutions such as rectors' conferences). To keep the web tool free of charges, especially for students, an equivalent to the charges could be paid by the EC.

Table 8-6: Funding scenario 2

COST FACTOR	COST SHARING
Basic fixed costs	50% principals 50% participation fees
Rankings	100% participation fees and user charges

The third scenario uses a market-oriented approach, trying to benefit as much as possible from a variety of market funding sources, minimizing the stable basic funding.

Table 8-7: Funding scenario 3

COST FACTOR	COST SHARING
Basic fixed costs	25% principals 25% fundraising 25% development projects 25% participation fees
Rankings	70% media partner, data providers, publishing companies 30% selling of products, user charges

The different scenarios could be seen as extreme cases, each of them focusing strongly on one or some of the potential funding sources. Looking at the different scenarios, the following general conclusions can be drawn:

- A completely market-oriented solution with basic lump sum funding near zero is extremely difficult to reach since U-Multirank with its related surveys is an expensive form of ranking and the commercial sources are limited.
- A comprehensive non-profit basic funding is also unrealistic; the logic of the funding structures of EC, foundations etc are not directed to large-scale and long-term basic funding.
- Free data provision, especially from nationally financed and run projects or from other existing data sources, will lower the cost of data collection. The more that national statistics offices harmonize data collection, the lower the costs will be.
- Charges to the users of the U-Multirank web tool would seriously undermine the aim of creating more transparency in European higher education by excluding students for example; but there is a possibility of some cross-subsidization from selling more sophisticated products such as data support to institutional benchmarking processes, special information services for employers, etc. The EC could pay for student user charges.
- Project-based funding for special projects, for instance new methodological developments or rankings of a particular 'type' of institution offer an interesting possibility with chances of cross-subsidization.
- Market revenues could come from commercial elements of the web tool (advertising, apps). As soon as it is possible to publish authoritative rankings publishers/media partners could contribute to the costs.

- A major additional source of income would be to charge institutional subscription fees (which could also be paid by governments or foundations on the institutions' behalf). For U-Map this seems to be a viable solution, but we are not sure if it would work for U-Multirank. The questions are: would paying for rankings produce a 'value for money' attitude on the part of institutions (or countries)? Would institutions be willing to pay for a product that entails a significant additional workload for them? What kind of benefit do institutions have to have in order for it to outweigh the costs of data gathering plus subscription fees?
- The resulting picture from the previous potential sources is a substantial level of basic funding (from three possible sources: EC, foundations, other sponsors) with a combination of a variety of market sources contributing cost coverage plus some cost reductions through efficiency gains.

8.9 A concluding perspective

U-Multirank is not only feasible in terms of methodology but also in terms of practical implementation. The implementation of the system will be a dynamic process over the coming years; after a next project phase of two years institutionalisation of a U-Multirank unit could be organized for the longer term. Within the project phase the instruments have to be finalized and developed in a user-friendly way and a business plan has to be designed. After two more years, the roll-out of the system should include about 700 European higher education institutions and about 500 institutions in the field-based ranking for each of three fields. The roll-out should cover Europe, including all institutional profiles, and should enlarge this database internationally, targeting the institutions required to reach sufficient coverage for all relevant profiles. The nature of the ranking has to remain global and should not merely serve European interests. Data from U-Multirank, U-Map, national field-based rankings and national statistics should be coordinated and integrated to allow 'pre-filling'. Despite the focus on the flexible web tool, concepts for authoritative rankings, either for the public or for associations of higher education institutions, should be developed because of their market potential.

Organizational options such as market, stakeholder, government or independent non-profit models should be seen as complementary approaches. But the organizational basis of U-Multirank should be the non-profit model with elements of the other options included. In particular, the aim of financial self-sustainability for U-Multirank makes the combination of some non-profit basic funding with the offer of commercial products inevitable. These two orientations should be reflected in the future organizational and governance structure of U-Multirank. The analysis of the fixed and flexible cost determinants could lead to a calculation of the cost, showing ambitious financial targets which are feasible if funded through a diversified funding base.

9 List of References

Athiyaman, A. (1997). Linking student satisfaction and service quality perceptions: the case of university education. European Journal of Marketing, 31(7), 528-540.

AUBR Expert Group. (2009). Assessing Europe's University-Based Research -- Draft. s.l. [Brussels]: European Commission – DG Research.

Azagra-Caro, J.M., de Lucio, I.F., & Gutierrez, G.A. (2003), ''University patents: Output and input indicators of what?'' Research Evaluation, 12(2): 5–16.

Balconi, M., Breschi, S., & Lissoni, F. (2004), ''Networks of inventors and the role of academia: an exploration of Italian patent data''. Research Policy, 33: 127–145.

Becher, T. and M. Kogan (1992). *Process and structure in higher education*. London, Heinemann.

Brandenburg, U & Federkeil, G. (2007) *How to measure internationality and internationalisation of higher education institutions! Indicators and key figures,* CHE Working paper No. 92

Brown, R. M., & Mazzarol, T. W. (2009). The importance of institutional image to student satisfaction and loyalty within higher education. *Higher Education, 58*(1), 81-95.

Bucciarelli, L.L. (1994), *Designing Engineers*, Cambridge, MA: MIT Press

CHERPA-Network. (2010). *Interim progress report: Design phase of the project 'Design and testing the feasibility of a multi-dimensional global university ranking'*. Enschede: CHEPS, University of Twente.

Clark, B. R. (1983). *The higher education system: academic organization in cross-national perspective*. Berkeley, University of California Press.

Cremonini, L., Benneworth, P., & Westerheijden, D. F. (2011). *In the shadow of celebrity: The impact of world class universities policies on national higher education systems*. Paper presented at the CHER 23rd annual conference.

Cremonini, L., Westerheijden, D. F., & Enders, J. (2008). Disseminating the Right Information to the Right Audience: Cultural Determinants in the Use (and Misuse) of Rankings. *Higher Education, 55*, 373-385.

Debackere, K., & Veugelers, R. (2005), The role of academic technology transfer organizations in improving industry-science links, *Research Policy* 34 (2005), pp. 321-342.

Dill, D. D., & Soo, M. (2005). Academic Quality, League Tables, and Public Policy: A Cross-National Analysis of University Ranking Systems. *Higher Education, 49*, 495-533.

Dulleck, U. and R. Kerschbamer (2006). "On Doctors, Mechanics, and Computer Specialists: The Economics of Credence Goods " *Journal of Economic Literature* **44**(1): 5-42.

Enquist, G. (2005) *The internationalisation of higher education in Sweden, the National Agency for Higher Education*, Högskoleverkets rapportserie 2005:27 R, Stockholm

Espeland, W. N., & Saunder, M. (2007). Rankings and Reactivity: How Public Measures Recreate Social Worlds. *American Journal of Sociology, 113*(1), 1-40.

Eurostat (2011), 'Patent Statistics at Eurostat: Methods for Regionalisation, Sector Allocation and Name Harmonisation', *Methodologies & Working papers*, Publications Office of the European Union, Luxembourg, 2011, ISBN 978-92-79-20237-7.

Expert Group on Knowledge Transfer Metrics (2009), Metrics for Knowledge Transfer from Public Research Organizations in Europe. Brussels: European Commission DG for Research. See: http://ec.europa.eu/invest-in-research/pdf/download_en/knowledge_transfer_web.pdf.

Federkeil, G. (2009). Reputation indicators in rankings of higher education institutions. In B. Kehm & B. Stensaker (Eds.), *University Rankings, Diversity, and The New Landscape of Higher Education* (pp. 19-34). Rotterdam; Taipeh: Sense Publishers.

Furco, A. & Miller, W. (2009), Issues in Benchmarking and Assessing Institutional Engagement, *New Directions for Higher Education*, No. 147, Fall 2009, p. 47-54.

Hazelkorn, E. (2011). *Rankings and the Reshaping of Higher Education: The Battle for World-Class Excellence*. London: Palgrave Macmillan.

Heine, C., & Willich, J. (2006). Informationsverhalten und Entscheidungsfindung bei der Studien- und Ausbildungswahl Studienberechtigte 2005 ein halbes Jahr vor dem Erwerb der Hochschulreife. . *HIS: Forum Hochschule*(3).

Holi M.T., Wickramasinghe, R. and van Leeuwen, M. (2008), *Metrics for the evaluation of knowledge transfer activities at universities*. Cambridge: Library House.

Hollander, E. L., Saltmarsh, J., and Zlotkowski, E., (2001), "Indicators of Engagement." In L. A. Simon, M. Kenny, K. Brabeck, & R. M. Lerner (eds.), *Learning to Serve: Promoting Civil Society Through Service-Learning*. Norwell, Mass.: Kluwer.

IAU, International Association of Universities (2005). *Global Survey Report, Internationalization of Higher Education: New Directions, New Challenges*, Paris: IAU.

International Ranking Expert Group. (2006). Berlin Principles on Ranking of Higher Education Institutions. Retrieved 24.6, 2006, from http://www.che.de/downloads/Berlin_Principles_IREG_534.pdf

Ischinger, B. and Puukka, J. (2009), Universities for Cities and Regions: Lessons from the OECD Reviews, *Change: The Magazine of Higher Learning,* Vol. 41, No. 3, p.8-13.

Iversen, E.J., Gulbrandsen, M., & Klitkou, A. (2007), "A baseline for the impact of academic patenting legislation in Norway". Scientometrics, 70(2), 393–414.

Jaschik, S. (2007, 03-19). Should U.S. News Make Presidents Rich? *Insude Higher Education*. Retrieved from http://www.insidehighered.com/news/2007/03/19/usnews

King, Gary et al (2004): Enhancing the Validity and Cross-cultural Comparability of Measurement in Survey Research. *American Political Science Review* 98, p. 191 -207.

King, Gray, Jonathan Ward: Comparing Incomparable Survey Response: Evaluating and Selecting Anchoring Vignettes. *Political Analysis* 15, p. 46 -66

Lissoni, F., Llerena, P., McKelvey, M. and Sanditov, B. (2008), ''Academic patenting in Europe: New evidence from the KEINS database''. Research Evaluation, 17(2): 87-102.

Magerman T, Grouwels J., Song X. & Van Looy B. (2009). Data Production Methods for Harmonized Patent Indicators: Patentee Name Harmonization. EUROSTAT Working Paper and Studies, Luxembourg.

Marginson, S. (2006, 7-9 September). *Global university rankings: private and public goods.* Paper presented at the 19th Annual CHER conference, Kassel

McDonough, P. M., Antonio, A. L., & Perez, L. X. (1998). College rankings: democratized college knowledge for whom? . *Research in Higher Education, 39*(5), 513-537.

Meyer, M., Sinilainen, T. and Utecht, J.T. (2003), ''Toward hybrid triple helix indicators: A study of university-related patents and a survey of academic inventors''. Scientometrics, 58: 321–350.

Montesinos; P., Carot; J.M., Martinez; J.M., Mora, F. (2008), Third Mission Ranking for World Class Universities: Beyond Teaching and Research, *Higher Education in Europe*, Vol. 33, Nr. 2, pp. 259-271.

Nelson, P. (1970). "Information and consumer behavior." *The Journal of Political Economy* **78**(2): 311-329.

Nuffic (2010) *Mapping internationalization,* http://www.nuffic.nl/international-organizations/services/quality-assurance-and-internationalization/mapping-internationalization-mint

OECD (2003), *Turning Science into Business: Patenting and licensing at public research organizations*. Paris: OECD.Rojas-Méndez, J. I., Vasquez-Parraga, A. Z., Kara, A., &

Cerda-Urrutia, A. (2009). Determinants of Student Loyalty in Higher Education: A Tested Relationship Approach in Latin America. *Latin American Business Review, 10*(1), 21-39.

Oudshoorn, N. & Pinch, T. (Eds.).(2003), *How Users Matter, the Co-construction of Users and Technology*, Cambridge, MA: MIT Press

Peeters B., Song X., Callaert J., Grouwels J., Van Looy B. (2009). Harmonizing harmonized patentee names: an exploratory assessment of top patentees. EUROSTAT working paper and Studies, Luxembourg.

Popper, K. R. (1980). *The Logic of Scientific Discovery* (rev. ed.). London: Hutchinson.

Rauhvargers, A. (2011). *Global University Rankings and their Impact*. Brussels: European University Association.

Sadlak, J., & Liu, N. C. (Eds.). (2007). *The world-class university and ranking: Aiming beyond status*. Bucharest; Shanghai; Cluj-Napoca.

Salmi, J. (2009). *The Challenge of Establishing World-Class Universities*. Washington, D.C.: World Bank.

Saragossi, S., & van Pottelsberghe, B. (2003), ''What patent data reveal about universities: The case of Belgium''. Journal of Technology Transfer, 28: 47–51.

Schmiemann, M. and Durvy, J.-N. (2003), New approaches to technology transfer from publicly funded research in: *The Journal of Technology Transfer*, vol. 28, pp. 9-15.

Teichler, U. (2004), The changing debate on internationalisation of higher education, *Higher Education* 48 (1), 5-26.

The CHE Ranking of European Universities: A Pilot Study in Flanders and the Netherlands. (2008). s.l. [Gütersloh, Enschede, Leiden, Brussels]: CHE; CHEPS; CWTS Leiden University; Vlaamse Overheido. Document Number)

Thibaud, A. (2009). *Vers quel classement européen des universités ? Etude comparative du classement de Shanghai et des autres classements internationaux* (No. Note de Benchmarking 4). Brussels; Paris: Institut Thomas More.

Thursby, .J, Fuller, A., & Thursby, M. (2007), "US faculty patenting: Inside and outside the university." *NBER Working Paper 13256*. Cambridge, MA: National Bureau of Economic Research.

Tijssen, R. F. W. (2003). Scoreboards of research excellence. *Research Evaluation, 12*(2), 91-104.

Tijssen, R.J.W., T.N van Leeuwen, and E. van Wijk, (2009) Benchmarking university-industry research cooperation worldwide: performance measurements and indicators based on co-authorship data for the world's largest universities, Research Evaluation, 18, pp. 13-24.

Tijssen, R.J.W., Waltman, L., and N.J. Van Eck (2011) Collaborations span 1,553 kilometres, Nature, 473, p. 154.

Usher, A., & Savino, M. (2006). *A world of difference: A global survey of university league tables*. Toronto: Educational Policy Institute.

Van Dyke, N. (2005). Twenty Years of University Report Cards. *Higher Education in Europe, 30*(2), 103-125.

Van Raan, Anthony (2003): Challenges in the Ranking of Universities. In: Jan Sadlak, Liu Nian Cai (eds.): *The World-Class University and Ranking: Aiming Beyond Status.* Bucharest: UNESCO-CEPES.

van Vught, F. A. (2008). Mission diversity and reputation in higher education. *Higher Education Policy 21*(2), 151-174.

van Vught, F. A., Kaiser, F., File, J. M., Gaethgens, C., Peter, R., & Westerheijden, D. F. (2010). *U-Map: The European Classification of Higher Education Institutions*. Enschede: CHEPS.

Waltman, L., R.J.W. Tijssen, and N.J van Eck (2011) Globalisation of science in kilometres, Journal of Informetrics

Westerheijden, D. F., Beerkens, E., Cremonini, L., Huisman, J., Kehm, B., Kovac, A., et al. (2010). *The first decade of working on the European Higher Education Area: The Bologna Process Independent Assessment - Volume 1- Detailed assessment report*. s.l. [Brussels]: European Commission, Directorate-General for Education and Culture. Document Number)

Yarbrough, D. B., Shulha, L. M., Hopson, R. K., and Caruthers, F. A. (2011). *The program evaluation standards: A guide for evaluators and evaluation users* (3rd ed.). Thousand Oaks, CA: Sage