EUROPEAN COMMISSION / European Research Area / Science & society

Assessing Europe's University-Based Research

Expert Group on Assessment of University-Based Research

Research Policy

EUR 24187 EN

Interested in European research?

Research*eu is our monthly magazine keeping you in touch with main developments (results, programmes, events, etc.).
It is available in English, French, German and Spanish. A free sample copy or free subscription can be obtained from:

European Commission
Directorate-General for Research
Communication Unit
B-1049 Brussels
Fax (32-2) 29-58220
E-mail: research-eu@ec.europa.eu
Internet: http://ec.europa.eu/research/research-eu

EUROPEAN COMMISSION

Directorate-General for Research
Directorate C - European Research Area: Knowledge-Based Economy
Unit C.4 - Universities and Researchers

Contact: Adeline Kroll

European Commission
Office SDME 9/17
B-1049 Brussels

Tel. (32-2) 29-85812
Fax (32-2) 29-64287
E-mail: Adeline.Kroll@ec.europa.eu

EUROPEAN COMMISSION

Assessing Europe's University-Based Research

Expert Group on Assessment of University-Based Research

RTD.C4

Directorate-General for Research

2010 Science in Society 2008 Capacities, 1.4.1 EUR 24187 EN

EUROPE DIRECT is a service to help you find answers to your questions about the European Union

Freephone number (*):

00 800 6 7 8 9 10 11

(*) Certain mobile telephone operators do not allow access to 00 800 numbers or these calls may be billed

LEGAL NOTICE:

Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use which might be made of the following information.

The views expressed in this publication are the sole responsibility of the author and do not necessarily reflect the views of the European Commission.

A great deal of additional information on the European Union is available on the Internet. It can be accessed through the Europa server (http://europa.eu).

Cataloguing data can be found at the end of this publication.

Luxembourg: Publications Office of the European Union, 2010

ISBN 978-92-79-14225-3
ISSN 1018-5593
doi 10.2777/80193

Printed in Belgium

PRINTED ON WHITE CHLORINE-FREE PAPER

FOREWORD

How to create a new and more coherent methodology to assess the research produced by European universities?" This is the question experts were asked to answer, following a 2006 Commission Communication on the modernisation of universities1, which suggested that universities should become more specialised and concentrate on working to their specific strengths.

Universities rankings are increasingly popular. Today, 33 countries have some form of ranking system operated by government and accreditation agencies, higher education, research and commercial organisations, or the media. The most popular are the Shangai Jiao Tong Academic Ranking of World Universities and the Times QS World University Ranking.

Rankings are used for specific and different purposes. Politicians regularly refer to them as a measurement of their nation's economic strength and aspirations. Universities use them to define performance targets and implement marketing activities, while academics use rankings to support their own professional reputation and status. Students use rankings to choose their potential place of study and research. Public and private stakeholders use rankings to guide their decisions about funding allocations. What started out as a consumer product aimed at undergraduate domestic students has now become both a manifestation and a driver of global competition and a battle for excellence in itself.

However while there are over 17,000 higher education institutions worldwide, rankings concentrate interest only in the world's top 100.

In addition, if higher education is one of the engines of the economy and a key point on the 'knowledge triangle', then the productivity, quality and status of research produced by universities is a vital indicator. Hence the importance of designing a way to evaluate it which is truly fit for purpose. But, as always, there is no 'one size fits all' solution! A new methodology will have to be developed. Ideally the best would be applicable across a full range of disciplines, including interdisciplinary research. It should assume an inclusive notion of research, ranging from blue sky/curiosity-driven to user-led/practice-based research. At present, some rankings include metrics on teaching and learning, most are focused on life-science research.

Users too have their own specific needs. And, depending on what they want to find out, they should be provided with a broad range of answers. For example, a prospective student might look for information on a specific discipline, on future employability, or on the fees associated with the university of their choice. A ranking system of this kind does exist for students, but at the moment only in Germany. The level at which the quality of research is assessed also matters. Ranking universities as entire institutions may not be the most appropriate way to identify where the best research is done and how it is done. A university may be renowned for one or two departments, but may not be excellent in all disciplines it offers. Identifying more precisely where research is produced and disseminated should allow for a better assessment of university-based research.

I believe that the coexistence of different models to assess university-based research is not only inevitable, but healthy. We need to design flexible and multidimensional methodologies that will adapt to the diverse and complex nature of research, disciplines and of our universities. In its quest for excellence, the European Commission must and will encourage, promote and support every effort to understand and monitor the quality of research at universities.

I wish to end with a simple quote from someone who understood better than anyone else the value of freedom, creativity and knowledge: “Not everything that counts can be counted, and not everything that can be counted counts.” This sign was hanging in Einstein’s office at Princeton. Let us take the time now to see what really counts when we rank our universities, these most important of our knowledge powerhouses.

Commissioner Janez POTOČNIK

Overview

1.1 Executive Summary

HEIGHTENED IMPORTANCE OF UNIVERSITY-BASED RESEARCH AND OF ASSESSMENT OF UNIVERSITY-BASED RESEARCH

The political context

Assessment of university-based research[1] (AUBR) has become a major issue for a wide range of stakeholders at all levels. One of the main reasons is that research performance is widely regarded as being a major factor in economic performance. Because of their interlinked roles in education, research, and innovation, universities are considered key to the success of the Lisbon Strategy with its move towards a global and knowledge-based economy. Improving the capacity and quality of university-based research is thought to be vitally important for innovation, including social innovation. In the words of the revised Lisbon Strategy (European Commission (2005), p. 20), "knowledge, meaning R&D, innovation and education, is a key driver of productivity growth. Knowledge is a critical factor with which Europe can ensure competitiveness in a global world". According to the Commission (p. 20), the economic relevance of research requires, among other things, 'increased and more effective public expenditure', a view that is shared by an increasingly large number of Member States.

The economic dimension of (university-based) research in terms of expected economic and societal benefit and increased expenditure goes a long way to explain the heightened concern for quality and excellence in research, for transparency, accountability, comparability and competition, and for performance indicators and assessment. The following quote from the Commission's Communication *Delivering on the modernisation agenda for universities: Education, research and innovation* of 2006 (p. 7f.) illustrates this:

> Universities should be funded more for what they do than for what they are, by focusing funding on relevant outputs rather than inputs, … Competitive funding should be based on institutional evaluation systems and on diversified performance indicators with clearly defined targets and indicators supported by international benchmarking.

Global rankings

The growing concern for the quality and assessment of university-based research partly explains the increasing importance attached to university rankings, especially global rankings. As is well known, rankings compare universities on the basis of a range of indicators; different systems favour different indicators, and the same indicators can be weighted differently by the various systems. The total score for each university is aggregated into a single digit, and universities are ranked accordingly. Rankings enjoy a high level of acceptance among stakeholders and the wider public because of their simplicity and consumer-type information. However, assessment experts have expressed serious reservations about the methodologies used by global ranking organisations. In particular, doubt has been cast on the possibility of comparing whole universities – in other words,

[1] In this report, the term 'university' refers to all higher education institutions (HEIs), irrespective of the name and status in national law.

diverse and complex organisations – on the basis of aggregated scores. Moreover, global rankings tend to rely on qualitative indicator-based data, which tend to have an inbuilt bias in favour of hard sciences and biosciences, and of English-language publications. There is also a substantial lack of cross-national comparative data.

THE RAISON D'ÊTRE OF THE ASSESSMENT OF UNIVERSITY-BASED RESEARCH EXPERT GROUP

In this context, the Commission's Directorate-General for Research decided to convene an expert group on assessment of university-based research. On the one hand, there is the generally recognised need for performance assessment of university-based research, especially publicly funded research; on the other hand, there is a host of ranking and assessment systems, few of which – if any – seem to do justice to the diversity of research disciplines and fields, of research outputs, of university profiles, and – by no means least important – of users and user needs and interests.

Assessment of University-based Research Expert Group: remit and composition

The Expert Group on Assessment of University-based Research was established in July 2008 to identify the parameters to be observed in research assessment as well as analyse major assessment and ranking systems with a view to proposing a more valid comprehensive methodological approach. The overall objective was to promote and contribute to the development of multidimensional methodologies designed to facilitate the assessment of university-based research. Specifically, the Group was asked to

- Identify the various types of users (or potential users) of measurements of the quality of university-based research, and to analyse their purposes and needs;
- Take stock of the main methodologies for assessing / ranking the quality of university-based research, including existing international assessments / rankings and other methodologies being developed, with a view to understanding their purpose, scope, uses, merits, limitations, biases and impact;
- Propose, as far as possible, a consolidated multidimensional methodological approach addressing the various user needs, interests and purposes, and identifying data and indicator requirements.

The Expert Group had 15 members from 12 EU Member States, Australia, a European association and an international organisation. Members represented a wide range of pertinent backgrounds, including experience and/or expertise in national and international rankings and bibliometrics, data collection and analysis, concrete research assessment exercises, the workings of leading national and European research funding organisations, collaboration with OECD, participation in pertinent EU expert groups and projects, and university senior management. Academically speaking, the experts represented a variety of disciplines, including arts and design, humanities, socio-economic sciences, and natural sciences.[2]

That people from such diverse backgrounds with initially different views on, inter alia, assessment methods and appropriate research outputs and outcomes reached agreement on a number of basic principles and a new approach to AUBR, plus a number of action-

[2] See Appendix 1 for the CVs of Group members.

oriented recommendations, lends credibility to these principles and recommendations and the approach proposed. Members of the Group would like to believe that reaching consensus on the key issues to be observed in AUBR and on how to address them constitutes a major achievement in its own right.

Activities undertaken and outcomes reached by the AUBR Expert Group

Two major interrelated activities were undertaken: i) preparation and discussion of a number of comprehensive topic-specific working papers, and ii) preparation and analysis of case studies of institutional, national and global assessment exercises and systems. Moreover, the Group conducted a workshop in order to have the provisional outcomes of their work validated by invited key experts and stakeholder representatives.

(i) Topic-specific working papers

Activities

Working papers on the following topics and issues were prepared:

- **Users and uses:** Who are the expected end users – target users, anticipated users, and accidental users - of the multidimensional tool envisaged, and of the results of assessments undertaken, and what are their purposes, needs, requirements, and interests?
- **Methods:** What are the main methods used for assessing research, and what are their characteristics, uses, merits, limitations, biases, and impact?
- **Disciplines:** Which disciplines should be considered in assessment exercises? How is / should research in the various disciplines, and how should interdisciplinary research be assessed? To what extent is or should the issue of languages be considered in research assessment?
- **Research:** How is research defined by major international and national organisations? How far along the Research-Development-Innovation spectrum should activities and outputs be included?
- **Social impact:** Bearing in mind the Lisbon Agenda, how is / can social and economic relevance, benefit and impact be measured?
- In a second step, the Group further developed and synthesised these working papers. In addition to the paper on users and uses, members of the Group prepared working papers on **Research and disciplines** and **Indicators and impact**.

Outcomes

(1) Comprehensive overview of users and uses

A comprehensive survey of stakeholders and their requirements, prepared as part of the working paper on users and uses, gave rise to the following conclusions.

- Individual user groups have a wide range of AUBR-related requirements;
- While there is overlap between the requirements of different user groups, some needs are specific to particular groups;

- While some of the data required may be readily available or relatively easy to obtain, other data are either not available or only available in limited circumstances. This makes comparability across universities and countries difficult.

(2) Basic principles to be observed in the assessment of university-based research.

Guided by the conviction that the purpose and objectives of a given assessment exercise should be seen in context, the Group identified a number of basic principles. In particular, assessment of university-based research should

- Cover all disciplines and, crucially, trans-, multi-, and interdisciplinary work, as well as research in emerging new disciplines;
- Recognise the whole spectrum of research, including fundamental, applied, and practice-based research;
- Take into account the diverse outputs and outlets through which research outcomes are transmitted and disseminated;
- Recognise the different dimensions of research, namely input, process, output, and outcome;
- Take account of social, economic, environmental, and cultural impact and benefits;
- Be cognisant of the diverse profiles, missions, histories, and resources of Europe's universities and higher education systems;
- Take into consideration, as appropriate, research tailored to specific local, regional and national contexts and / or published in languages other than English;
- Include research carried out in bilateral partnerships, in partnerships with non-academic organisations, and in European and international networks.

(3) Strengths and weaknesses of the various indicators used in assessment exercises

The Expert Group analysed the different characteristics and dimensions of indicators, and their advantages and disadvantages. It studied both the value and limitations of bibliometric data which are commonly used to measure research productivity and quality, and scientific-scholarly impact. In accordance with observations on the requirements of different users, and the basic principles proposed, the following conclusions were reached:

- There is no single set of indicators capable of capturing the complexity of research and research assessment.
- There is no such thing as a perfect indicator; all indicators have their own specific strengths and weaknesses, and assessment exercises have to take this into consideration from the outset.
- There is no such thing as an objective indicator: Indicators are rarely a direct measurement; more often than not, they are proxies.
- Indicators must be fit for purpose and verifiable.
- The different publication and dissemination practices characteristic of different disciplines and fields can be positively and negatively affected by the choice of indicators. This is also true for bibliometric accounting, which currently tends to favour specific disciplines to the detriment of others.

Hence, the choice, interpretation, and weighting of indicators are of utmost importance in any assessment exercise or system.

The report provides, in table format, a comprehensive overview of the most commonly used indicators, relating each indicator to the measurement of a specific aspect or dimension of research, pointing out its strengths and weaknesses, and indicating some further development which should be undertaken to make the indicator in question more robust.

In this context, it is worth mentioning that the onset of the global financial and economic crisis, which occurred during the course of the Expert Group's deliberations, had an impact on the Group's discussions. For example, the Group came to realise that even bibliometric indicators might be flawed due to manipulation of data.

(ii) Case studies

Activity

To obtain a clearer idea of existing methodologies for assessing university-based research and for ranking and rating universities/units within universities, case studies of pertinent exercises were prepared. Members of the Expert Group reviewed practices in their home countries and universities.

Outcomes

These case studies represent different approaches and objectives. They furnish evidence that universities and national organisations regard assessment of university-based research as important for improving research performance and quality, for strategic planning and for international benchmarking. They also reveal the common view that global rankings are not the perfect answer to their requirements. The case studies highlight a number of key aspects of assessment, which confirm or complement the insights gained and the principles developed by the Expert Group. The following are particularly noteworthy.

- **Consultation of HE researchers** in the development of assessment systems to ensure procedural fairness, transparency, and a high level of acceptance;
- The use of **peer review panels,** to ensure a broader understanding of the research being assessed, as well as of its contribution to knowledge, and to facilitate the assessment of research in emerging new disciplines and of interdisciplinary research;
- The **combination** of **peer assessment** and **bibliometric indicators**;
- The use of information about **process** and **impact,** including **impact on teaching,** to balance the focus on research output;
- **Self-evaluation** as a key component in the assessment process;
- Experiments designed to facilitate the **measuring of societal impact**;
- Focus on **units of assessment** positioned somewhere between the individual researcher and the entire institution;
- **Unintended consequences** of assessment exercises, be it that stakeholders make decisions contrary to the original objective(s) pursued, or be it that research quality is made the focus of attention to the detriment of other university functions.

OVERARCHING OUTCOME: FAIRNESS AND FEASIBILITY IN ASSESSMENT OF UNIVERSITY-BASED RESEARCH – A NEW APPROACH

On the basis of the principles and insights gained from the preparation and discussion of the working papers, and of the key aspects identified in the case studies, the Expert Group wishes to make the following **general recommendations** and, linked to these, propose **a new approach to AUBR**.

General recommendations

(1) Assessment of university-based research should be designed in relation to purpose and articulated objectives, and employ methodologies that are fit for purpose.

(2) Assessment of university-based research should combine quantitative indicator-based data with qualitative information, for example information based on expert peer assessment or validation, or/and end-user review.

(3) Assessment of university-based research should be undertaken at the level of 'knowledge clusters', the precise scale and nature of which depends on the purpose of the research exercise. Knowledge clusters may be based on administrative units (e.g. faculties, departments, schools, teams, centres, institutes, interdisciplinary issue-driven clusters, etc.), fields of science within universities or inter-institutional networks. Knowledge clusters should allow for aggregation to institutional level.

A new approach

The Expert Group developed the outline of a multi-dimensional research assessment matrix. It links specified users with their defined purposes and objectives to specific data, quantitative and qualitative indicators, and specific assessment methods. While some purposes and objectives require extremely detailed and robust data on research outputs, other requirements demand only a few, relatively simple indicators.

Chapter 5 of the Report sketches a number of *if-then* scenarios, which illustrate the approach proposed. Among the purposes specified in the scenarios are allocation of resources, improvement of research performance, mission differentiation, and attraction of talent.

As user purposes and objectives frequently overlap, a comprehensive web-enabled and personalized tool-kit can be readily developed to meet different policy and university needs.

External validation of provisional outcomes

In April 2009, the Group organised, together with the European Commission, a workshop, which was opened by Commissioner Potočnik, and attended by some twenty external experts and representatives of stakeholder organisations, and 15 officials from DG Research and other Commission services. The key objective was to validate the provisional results of the Expert Group's work.

External workshop participants welcomed the principles and recommendations presented by EG members. They suggested that further consideration should be given to the concept of the 'knowledge cluster', especially with regard to the conflicting principles of 'diversity' and 'comparability', and to the combination of indicators and peer review. In addition, external participants made a number of recommendations for follow-on activities, regarding improvements in bibliometrics to cover all disciplines; the development of new methods to

capture societal and economic impact, inter-disciplinary and collaborative research and activities across the full research-innovation eco-system; improvements in institutional capacity to collect, maintain, analyse and disseminate standardised data, so as to enable inter-institutional and cross-national comparisons. It was also suggested that the challenges and opportunities presented by the report of the Expert Group should be taken up in the current discussions about the further development of the European Research Area.

These and other recommendations were incorporated into the Expert Group's final report, and specifically into the recommendations to stakeholders set out below.

RELEVANCE OF THE FINAL REPORT TO POLICY DEVELOPMENT AND THE DESIGN OF FUTURE RESEARCH ASSESSMENT EXERCISES

The report is designed as a guide for

- Users of the outcomes of assessments of university-based research, enabling them to form a judgment on the adequacy or otherwise of existing and future assessment exercises and systems;
- Decision-makers in higher education, encouraging them to reflect on the unintended consequences that decisions solely based on research assessment might have;
- Specialists engaged in assessment of university-based research, presenting them with a number of basic principles that need to inform assessment of university-based research, and providing them with the outlines of a matrix for user- and purpose-driven multidimensional research assessment.

1.2 The Way Forward – Recommendations

The AUBR EG believes that, generally speaking, assessment of university-based research is being hampered by a lack of reliable, comparable, and comprehensive data. In view of this, the AUBR EG recommends that the European Commission

•Take the lead in establishing a European Observatory for Assessment of University-based Research to identify and prioritise data requirements of a European Research Assessment Framework, as well as to further develop and disseminate guidelines for use by universities, national agencies, government, and other stakeholders, based on the principles outlined in this report;

•Invest in developing a shared information infrastructure for relevant data to be collected, maintained, analysed, and disseminated across the European Union;

•Launch a project for the development and piloting of a full-fledged flexible Multi-dimensional Research Assessment Matrix, along the lines sketched in this report, enabling diverse users and stakeholders to design fit-for-purpose assessment scenarios, methodologies, and instruments;

•Adapt the Multi-dimensional Research Assessment Matrix to web-based technologies in order to facilitate personalisation, thereby meeting different user requirements, and substantially enhancing the Matrix's capability and user-friendliness.

•Launch a project for the development and piloting of indicators designed to measure the social and economic impact of research in general, and of European/ international collaborative research in particular;

•Develop a financial model to cover the full cost of university-based research including the cost of assessment, which is now an integrated element of the research process.

The AUBR EG notes that global university rankings have become a popular means of gauging university-based research. The EG cautions against rankings or similar assessment systems which seek to compare whole universities on the basis of an aggregated score and which lack validation through expert peer assessment. Moreover, in the absence of comprehensive reliable and comparable cross-national data, rankings cannot be a valid tool to achieve the overarching aim of improving the quality of university-based research across the European Union.

2 Introduction

This chapter outlines the national, European and international context for the establishment of the Expert Group on the Assessment of University-based Research (AUBR). This includes an overview of European policy and the influence of global rankings. The remit of the Expert Group is explained, and a summary of its activities and findings is presented.

2.1 University-based Research in the Knowledge Economy

Around the world, knowledge is recognised as the critical factor for global competitiveness in the 21st century. According to this view, successful societies are those most able to exploit knowledge for competitive advantage and performance, and attract international talent, new business and investment. This requirement for a knowledge-based society is central to European Union and national government strategies for sustainable economic and social development, especially as a response to the global economic crisis.

Because university-based research is the primary arena for the production of new knowledge, higher education is an important focal point for European Union and national government policy-making. It plays a critical role in the research-innovation eco-system, providing human capital through education and training, attracting high-skilled talent and investment, actively engaging with the local and regional community through knowledge and technology transfer, and underpinning the global competitiveness of nations and regions. Thus, the status, quality and productivity of higher education have become a national strategic objective and an indicator of global competitiveness.

According to the European Council Resolution (2007), European universities are at the forefront of 'Europe's drive to create a knowledge-based society and economy and improve its competitiveness'. To meet these challenges, universities are undergoing profound change. Competition is intensifying between universities nationally and internationally, students are becoming more conscious of the value of their education and its impact on their career opportunities, and governments and other stakeholders are asking questions and requiring evidence of value-for-money.

Attention is shifting to mechanisms to assess and benchmark the quality and performance of university teaching and learning, and of research performance. In recent years, there has been a steady growth in methods to evaluate and assess the activity and outcomes of higher education, with particular emphasis on the assessment of university-based research (AUBR). Cross-national or worldwide comparisons of research performance and quality are an inevitable outcome of globalisation.

University rankings have become popular around the world because they appear to provide a simple method to gauge world class excellence and provide accountability. In response to the results of the Shanghai Jiao Tong *Academic Ranking of World Universities* or Times QS *World University Rankings* concern has been expressed that too few European universities are ranked among the world's top 50 or 100 universities. These factors have contributed to the growing importance attached to research assessment, and the search for tools which can help improve research, identify value-for-money and allocate resources.

Research assessment can play an important role in improving performance and quality, supporting institutional autonomy and strategic planning, differentiating research missions and attracting talent. But indicators are not value-free. Measuring the wrong things can

easily distort. In order to capture the full richness of university-based research a multidimensional approach, combining qualitative and quantitative methodologies, is necessary.

According to Einstein: 'Not everything that counts can be counted, and not everything that can be counted counts.'

2.2 The European Policy Context

The Lisbon Agenda

In March 2000, the European Council (2000, section 5) agreed a new strategic goal to make Europe 'the most competitive and dynamic knowledge-based economy in the world, capable of sustainable economic growth with more and better jobs and greater social cohesion.' Because of the importance of research and development to 'generating economic growth, employment and social cohesion' the European Union (section 12) confirmed its support for the objectives of the European Research Area. In 2005, the European Council (p. 4) reaffirmed its commitment to increasing investment in R&D to 3% GDP in addition to as well as 'more numerous well-trained and motivated researchers'. *Working together for growth and jobs: A new Start for the Lisbon Strategy* (European Commission 2005, p. 9) says that European universities 'must be able to compete with the best in the world through the completion of the European Higher Education Area'.

Key policy documents

For realising the Lisbon Agenda, the European Commission and the Council of the European Union respectively set out concrete policies and strategies for bringing about the changes deemed necessary.

In its Communication 'Delivering on the Modernisation Agenda for Universities: Education, Research and Innovation', the Commission (2006, p. 7) argued that 'Universities should be funded more for what they do than for what they are, by focusing funding on relevant outputs rather than inputs...'

In its resolution 'Modernising Universities for Europe's Competitiveness in a Global Knowledge Economy', the Council (2007, p. 3) expressed the view that the 'challenges posed by globalisation require that the European Higher Education Area and the European Research Area be fully open to the world and that Europe's universities aim to become worldwide competitive players'.

Both documents stress the relevance of university-based research to attaining the Lisbon goals. Universities should communicate the relevance of their research to society and their stakeholders, and respond to calls for greater transparency, accountability and comparability.

Competitive funding should be based on institutional evaluation systems and on diversified performance indicators with clearly defined targets and indicators supported by international benchmarking for both inputs and economic and societal outputs … (European Commission 2006, p. 8).

Assessment methodologies should recognise and support the diversity of European universities, which differ in their history and degree of involvement in research (some focus more on research than others), the nature of their research activities (scientific, technical/applied research), their links to potential users of the results of their research (other universities, SMEs and large enterprises), the geographical scope of their research partners, and their user-community (local, regional, national, international).

The aim is to position European universities and research to generate increased investment, attract researchers from inside and outside Europe, enhance the impact of university-based research on SMEs and regional innovation, and strengthen teams engaged in inter- and trans-disciplinary research and global research networks.

EU initiatives designed to enhance the quality of university-based research

In addition to these and related policy documents, the EU has launched a number of initiatives designed to support the implementation of the policies proposed. These include:

•*Classifying European Institutions of Higher Education,* a pilot project funded, beginning 2004, by DG Education and Culture, which aims to classify European universities according to a multidimensional methodology.

•*Mutual Learning on Approaches to Improve the Excellence of Research in Universities,* an expert group launched by the Scientific and Technical Research Committee (CREST) in 2007. The overall objective was to conduct a mutual learning exercise on the scope, objectives and measures of national policies to improve research excellence in universities, to learn more about the effect of these policies, to identify good practices, and to develop recommendations for improving the policies and their impact on research in universities.

•*European University Data Collection*, a project studying the feasibility of a sustainable European system of data collection on the activities and performance of the European higher education institutions in the areas of education, research and innovation.

•*European Multidimensional University Ranking System,* a pilot project funded by DG Education and Culture, aimed at mapping multiple excellences (e.g. teaching, innovation, community engagement and employability). It complements the Classification project. First results are expected to be available in the first half of 2011.

Over recent years, there has been a steady growth in the number of national and international systems of research evaluation and assessment (see Appendix IV). They are usually operated by accreditation or quality assurance agencies, research councils and/or funding councils. There are similar international initiatives: The European Association for Quality Assurance in Higher Education (ENQA) was established to disseminate information, experiences and good practices in the field of quality assurance (QA) in higher education to European QA agencies, public authorities and higher education institutions. The European University Association's (EUA) *Institutional Evaluation Programme* focuses on quality enhancement at institutional level; it emphasizes self-evaluation and peer-review as a tool of institutional strategic development and improvement. The OECD AHELO (Assessment of Higher Education Learning Outcomes) project aims to assess learning outcomes on an international scale by establishing the validity of measures which cross cultures and languages.

2.3 Measuring what Counts

Many people think that university rankings provide an easy solution to measuring research performance and quality nationally and internationally. However, the experience of rankings illustrates that they can promote a simplistic understanding of university-based research and its contribution to society and the economy. Rankings can misinform and mislead when used to influence decision-making by governments, universities and other stakeholders.

University rankings have become an increasing influence on the higher education landscape since *US News and World Report* began providing consumer-type information about US universities in 1983. Since then, national rankings have been created in over 45 countries by public media organisations, government agencies or independent organisations. They usually use a combination of public or institutional data and/or peer or student surveys. Weightings are assigned to the individual indicators, and the total score is aggregated into a single digit ranking.

The Shanghai Jiao Tong *Academic Ranking of World HEIs* was first published in 2003, and become immediately popular and frequently quoted. It was followed quickly by the Times QS *World University Ranking* (henceforth Times QS, 2004), *Webometrics* or the *Ranking Web of World Universities* (2004), the Taiwan *Ranking of Scientific Papers for World Universities* (henceforth Taiwan, 2007), and the Leiden Ranking by the Centre for Science and Technology Studies (CWTS, 2009).

Rankings compare higher education institutions (HEIs) using a range of indicators, which are weighted differently by each ranking system. Considerable concern has been raised about their over-reliance on international bibliometric and citation databases, e.g. Thomson-Reuters *World of Science* or Elsevier-Scopus. These tend to benefit the physical, life and medical sciences and disadvantage engineering, social sciences, humanities and arts disciplines. They have an in-built bias in favour of English language publications, and against nationally-focused research. Citations measure impact on academic knowledge but ignore the important role that universities have in knowledge and technology transfer.

Concern has also been raised about the use of peer or reputation-based surveys, which are often self-perpetuating. In other words, because an institution is known as being excellent in some aspect, it is considered excellent in everything it does. This is called the 'halo' effect. Reputational surveys are also susceptible to 'gaming' which occurs when respondents deliberately downgrade competitors or upgrade their assessment in order to influence the outcome.

A wider issue concerns whether it is possible to measure and compare whole institutions. At a time of growing diversity of university mission and providers, rankings use a common set of indicators and weightings to measure all universities. Because global rankings focus on research intensity, other aspects of higher education, such as teaching and learning, community engagement, and third mission and innovation are ignored. In addition, universities are complex organisations with strengths and weaknesses across various departments and activities. An aggregate score is unable to reflect this. Rankings measure excellence differently depending upon the indicators and the weightings used.

Despite these concerns, international evidence suggests rankings are having a positive and perverse influence on decision-making by a growing number of stakeholder groups. They are associated with efforts to drive up research performance and quality, and allocate

resources. They can aid strategic management and planning, institutional profiling, identification of peer institutions, improve data collection and increase participation in broader discussions about institutional success.

Unintended consequences can occur when indicators are taken in isolation and simple correlations are made. This may include over-concentrating on research, favouring particular disciplines or allocating resources and realigning priorities to match indicators. This has included efforts by governments and universities alike to reframe strategies and priorities, and make significant changes at the system and institutional level in order to achieve a better ranking.

2.4 Remit of Expert Group on Assessment of University-based Research

Terms of Reference

Performance assessment of university-based research is increasingly important, especially for publicly funded research. Political and societal support for university research can only be maintained by a system of quality assessment, performance enhancement and value-for-money. Sound, verifiable and comparable data is a necessary prerequisite for institutional autonomy and to enable European universities to manage strategically, effectively and efficiently. It also assists universities to advance their own modernisation agenda, taking into account specific European values and objectives.

In response, the European Commission established the *Expert Group on the Assessment of University-Based Research* to develop a multidimensional methodology to assess the quality of research produced in universities, with a European perspective, and taking into account the diversity of European universities performing research, research disciplines, and the wide range of users. The aim is to enable institutional benchmarking, improvement in quality, and comparative assessment of universities across Europe.

The Terms of Reference were:

1) Review the needs of various types of users of measurement of research quality at universities;

2) Review main methodologies for assessing/ranking research quality of universities, covering existing international assessments/rankings and other methodologies being developed;

3) Propose as far as possible a consolidated multidimensional methodological approach addressing various users' needs, identifying data and indicators requirements (if necessary propose different approaches for different types of users).

Activities Undertaken by the Expert Group

The Expert Group met on seven occasions between July 2008 and July 2009 at meetings in DG Research, 8, Square de Meeûs, Brussels. The ‘core group’, composed of the Chairperson, the two Rapporteurs and the European Commission staff responsible for this activity, met prior to each meeting to prepare the work of the expert group.

A workshop with Commissioner Potočnik, fifteen members of the Expert Group plus more than twenty invited key experts and fifteen officials from different parts of the EU Commission was organised in April 2009 to validate the analysis and approach of the Expert Group. A summary of the workshop's conclusions is contained in Appendix V.

The Expert Group undertook a range of activities to inform its deliberations. This included an examination and analysis of national and global research assessment and/or ranking system, issues and methodologies associated with existent research assessment practices, and their impact and influence on research, researchers and higher education. Various working papers were prepared on the following issues:

- Identification of the range of users and uses of rankings and research assessment exercises;
- Detailed analysis of the existing and proposed indicators;
- Assessment of existing research assessment and ranking practices, and their impact and influence on research and disciplines, researchers and higher education;
- Preparation of a new framework for assessment of university-based research.

Outcomes and Achievements

Global rankings have achieved a high level of international popularity because they appear to provide a simple approach to the cross-national comparison and benchmarking of university-based research. The Expert Group concludes, however, that contrary to providing an accurate and useful assessment of research, rankings provide a very selective and potentially distorted picture. Recognising that research assessment must always be fit-for-purpose, the Expert Group has

1) Illustrated the wide range of users and uses of research assessment information;
2) Focussed on the full scope of activity across the research-innovation spectrum to which all disciplines have an important contribution to make;
3) Analysed the different characteristics and dimensions of indicators, and their advantages and disadvantages; and
4) Advocated a *Multidimensional Research Assessment Matrix to* provide the basis for strategic decision-making by the European Commission, national governments and universities.

2.5 Format of the Report

This report is divided into four main sections, plus appendices. ***Chapter 3*** sets out a number of guiding principles which have informed and underpinned the work of the Expert Group on AUBR. ***Chapter 4*** examines key characteristics of research assessment, and illustrates the complexity of the issues to be considered in the development of any research assessment exercise. ***Chapter 5*** proposes a *Multidimensional Research Assessment Matrix* which addresses various users' needs and identifies data and indicators requirements within a policy context. The concluding ***Chapter 6*** identifies potential risks and unintentional consequences which can arise if simplistic interpretations of the data are made, and one-dimensional correlations are drawn between research assessment and policy choices.

The Appendices provide significant reference material to facilitate 'good practice':

1) Case studies of research assessment and ranking systems across Europe and worldwide with description of the policy context, methodology and unintended consequences (Appendix IV); and
2) Encyclopaedia of indicators including a description, and the advantages and disadvantages (III).

3 Characteristics of Research Assessment

This chapter sets out the basic principles which have informed and underpinned the Expert Group's approach to research assessment. This embraces an inclusive definition of research and disciplines, and a broad understanding of differences between discipline research practices. Drawing on the experience of rankings and existent research assessment exercises, there is an extensive survey of likely users and uses of research assessment processes and results.

3.1 Defining 'Research'

The progression from simple to complex knowledge has, over decades, been reflected in the emergence of new disciplines, methodologies and ways of thinking, transforming societies and the way in which knowledge is created and used. Research, or the pursuit of new knowledge, has traditionally been divided into two major functions, basic and applied. Today, the boundaries between basic and applied research are blurring, and more and more fundamental research is conducted in the context of application, both within and outside universities.

The translation of research findings or knowledge into new or improved products and services is increasingly seen as an integral part of the research process. Knowledge has also become democratized in the sense that more people are aware of the issues and are social actors in the application of knowledge. Use-inspired research can be of a basic or fundamental nature. Universities are the primary organization for this type of research.

These developments have generated an important discussion on the definition R&D, and its outputs, outcomes, impact and benefits.

- The OECD Frascati Manual (OECD, 2002) says R&D 'comprises 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 New Knowledge Production model, developed by Gibbons et al. (1992), says that whereas traditional knowledge production (Mode 1) was disciplinary or 'curiosity-oriented' usually conducted by individuals in secluded/semi-secluded environment, Mode 2 knowledge is 'socially robust' and interdisciplinary, created within the context of being useful for the resolution of specific problems.

- Daniel Stokes (1997) devised Pasteur's Quadrant to distinguish various types of research according to whether or not it is inspired by a quest for fundamental understanding, and whether or not there are considerations of use.

- The EU (2005, 24 final) aims to overcome fragmentation of the knowledge system by linking the three elements of the 'knowledge triangle' – education, research and innovation – to encompass the whole innovation chain or the research, innovation and commercialization eco-system. This concept underpins the new European Institute of Technology and Innovation (EIT).

Based on these new understandings of knowledge production, research is viewed as a continuum, involving the whole process of discovery and spanning the spectrum from curiosity-driven to user-led, from blue-sky to practice-based.

3.2 An Inclusive Approach to Disciplines

These developments demonstrate that simple distinctions between STEM (science, technology, engineering and medicine) and HASS (humanities, arts and social sciences) disciplines ignore the complex and rich contribution that traditional and new disciplines make to our understanding of social, economic, scientific and technological challenges.

- *Mono-disciplinary* research is conducted within the boundaries of a specific discipline, contributing primarily to the advancement of knowledge in that discipline;
- T*rans-* or *multi-disciplinarity* bring together two or more disciplines without integration; and
- *Inter--disciplinarity* blends the approaches of two or more disciplines often leading to the creation of a new discipline.

Today, it is widely recognised that the major 'grand challenges' of humankind are not bound by borders or discipline. Complex global problems require interdisciplinary, collaborative solutions and inter-locking innovation systems. The United States *Committee on Facilitating Interdisciplinary Research* (2004) stated:

> Interdisciplinary thinking is rapidly becoming an integral feature of research as a result of four powerful 'drivers': the inherent complexity of nature and society, the desire to explore problems and questions that are not confined to a single discipline, the need to solve societal problems, and the power of new technologies (p2).

3.3 Research Outlets and Outputs

Research carries with it a responsibility to disseminate and apply the results of research activity professionally and ethically. An essential characteristic of research is that it leads to publicly verifiable outcomes which are open to authentication and scrutiny by experts.

Differences between disciplines or research fields derive from their history and the way in which research is conducted. These distinctions can lead to different forms of expression and outlet, which can affect the type of quantitative data and qualitative analysis. Depending upon the university, scientific field or policy environment, some formats may be more important than others.

- *Research outlet* refers to the avenues in which an output appears, such as journal name, conference, book publisher, theatre, art gallery, etc.;
- *Research output* refers to the individual journal articles, conference publications, book chapters, artistic performances, films, etc.
- *Form of publication*. Journals are the primary publication channel for almost all disciplines, but their importance differs across research disciplines. In some fields books (monographs) play a major role, while book chapters or conference proceedings have a higher status in other fields.
- *Other output formats.* While traditionally research has been published as academic texts, the complexity of knowledge has led to a diverse range of output formats, inter alia, 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.

Table 1 identifies the primary form of communications for the main discipline groups. For example, while natural and life scientists write books, their primary outlet is peer-reviewed journal articles. Engineering scientists primarily publish in conference proceedings although they also publish in journals and design prototypes. Social scientists and humanists have a wide range of outputs of which books are important sources of communication, while the arts produce major art works, compositions and media productions. In summary, Table 1 illustrates the diversity of research outlets, and why the focus only on journal articles cannot do justice to the contribution that other disciplines make.

Table 1 Primary Form of Written Communications by Discipline Group

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

3.4 Users and Uses

The assessment of university-based research has become a topic of increasing public attention. Their results are often published in the national media, ranked in descending order which is sometimes referred to as a 'league table'. Recently, cross-national comparisons of research performance, published as global university rankings, have emerged. They are an inevitable outcome of globalisation and their popularity illustrates that there is a wide audience for information on research performance and quality.

Drawing on the experience of rankings and existent research assessment exercises, Table 2 below identifies the wide range of potential users of research assessment information, the type of information that they are likely to require, and the purpose for which they are likely to use the information.

'Users' include policymakers and government agencies, universities, public or private research organisations (PROs), researchers or graduate students, employers, civil society and the media. Each group uses information differently to satisfy diverse, and often conflicting, objectives.

The experience with rankings suggests that different audiences use the results on research performance and quality for various reasons, many of which may not have been previously

considered. The experience also indicates that it is not possible to control how people may use or interpret the data.

Likely Target Users, including:

•*HE Governance and Management*: These groups require a wide range of information to help and inform policy and planning, permit strategic positioning and research strategy development and management, provide investor confidence and value-for-money guarantees, and underpin quality assurance. HE management is likely to use the information for publicity purposes – help student and academic recruitment, aid research partnerships (with other academic institutions or with public/private organizations), and initiate/sustain philanthropic relationships. Each group is likely to analyze the information differently. Because higher education is both a generator and user of the data, its position is different than the other users.

- ○ Governing Bodies or Councils
- ○ HE Executives and Management
- ○ HE Research Groups

•*Governments* – In addition to the reasons stated above, the groups below also require information to help define policy, to improve performance and quality and hence international competitiveness, to inform decisions about the higher education structure and the role of individual institutions, to ensure the HE system functions effectively and efficiently and in line with other government objectives. Local and regional authorities are likely to be interested in the reputation of individual universities and of the system as part of a wider economic strategy to position the city or region as an important node in the global economy. Capital and employment flows to where talent and quality education resides. If local/regional governments are a financial contributor to higher education, they will be interested in issues of value-for-money and efficiency, etc.

- ○ EU and Member Governments
- ○ Ministries of Education/Higher Education or Enterprise and Employment
- ○ Local and Regional Authorities
- ○ HE Agencies

•*Other Government Agencies* require good comparative data to assess the quality of research and HE performance and output and to support return-on-investment. Many research agencies also use the results of research assessment exercises to help aid resource allocation, while QA agencies use institutional data to benchmark and assess quality and performance.

- ○ Funding Agencies
- ○ Enterprise and Development Agencies

•*Academic Organisations and Academies* – In many countries, independent research and academic organisations, including academies, devoted only to research are as important as higher education institutions. Not only do they conduct research but they are also the primary professional academic body responsible for the 'accreditation' and valorisation of scientific quality.

•*Individuals* – Graduate students, especially international PhD students, are increasingly a keen user of cross-national comparisons, including rankings. They use the information to help inform their choice about the best institution to attend, including trends in graduate employment, which in turn is used as a proxy for career opportunities. Likewise, academics and researchers, including post-doctoral fellows, use this information to inform their decisions as to whether the institution in question is a place which values research and the quality and hence value of that research as perceived by others.

- Academics and Researchers (including post-doctoral fellows and visiting scholars)
- Students – most likely PhD students

Other Possible Users and Uses, includes:

•*Peer HEIs* – As international partnerships grow in prominence and strategic importance, HEIs are turning to rankings and other comparative information to help identify appropriate academic and research partners. But HEIs also use the information for strategic development purposes, helping to benchmark performance against appropriate peer institutions worldwide in order to learn and share experience, and improve performance.

•*Industry and other partner organizations* use HE performance indicators to help identify potential partners for projects, consultancy and knowledge/technology transfer. Increasingly, employers use such data to identify likely sources of potential employees.

- Private firms and entrepreneurs
- Public organizations
- Employers

•*Civil Society and Civic Organizations,* including non-governmental organizations (NGOs), are increasingly becoming partners and collaborators in research. This may include community and residential organizations, trade unions, etc.

- *Non-governmental organizations*
- *Community Organizations and Trade Unions*

•*Sponsors and Private Investors* – As HE diversifies its income sources, the role of benefactors, sponsors, philanthropy and private giving grows. These groups, which include alumni, are likely to use benchmarking data to identify potential 'investment' opportunities, using the information as a proxy for value-for-money and return-on-investment.

- Benefactors/Philanthropists
- Alumni

•*Ministries of Higher Education in Developing Countries* – A growing number of countries use rankings to award scholarships for overseas studies and determine which foreign higher education institutions are applicable.

Casual Users

•*Public Opinion*

The media is both a producer and an omnipresent communicator and transmitter of ranking-type information. In the first instance, the public has shown itself interested purchasing such information. But the media also performs an important public information role, helping ensure the public has a better and informed understanding of HE, its contribution and requirements (including financial), and providing investor (tax-payer) confidence. For higher education, the better informed the public is about such issues, the more likely the public will be to support higher education.

Table 2 Summary of User Groups and Uses of Research Assessment Data

User Group	Why Research Assessment Data is Required?	What Research Assessment Data is Required?
HE MANAGEMENT AND GOVERNANCE		
• Governing Bodies/Councils	• Policy and planning • Strategic positioning • Research strategy development/management • Investor confidence/value-for-money and efficiency • Quality assurance	• Institutional and discipline/field data re. level of intensity, expertise, quality and competence • Benchmarking against peer institutions, nationally and worldwide • Efficiency level: how much output vis-á-vis funding • Quality of academic staff and PhD students • Attraction capacity: recruitment of students, academics and researchers from outside region and internationally
• HE Executives/Management	• Policy and planning • Strategic positioning • Research strategy development/management • Investor confidence/value-for-money and efficiency • Quality assurance • Publicity • Student and academic recruitment • Improve and benchmark performance and quality	• Institutional and discipline/field data re. level of intensity, expertise, quality and competence • Benchmarking against peer institutions, nationally and worldwide • Efficiency level: how much output vis-a-vis funding • Quality of academic staff and PhD students • Attraction capacity: recruitment of students, academics and researchers from outside region and internationally • Identification of Partnerships (academic, public/private sector, NGOs, research organisations, etc.)
• HE Research Groups	• Strategic positioning • Research strategy development/management • Investor confidence/value-for-money and efficiency • Student and academic recruitment	• Discipline data re. level of intensity, expertise, quality and competence benchmarked against peer institutions • Quality of academic staff and PhD students • Attraction capacity: recruitment of students, academics and researchers from outside region and internationally • Identification of Partnerships (academic, public/private sector, NGOs, research organisations, etc.)
GOVERNMENTS AND GOVERNMENT AGENCIES		
• EU and National Governments	• Define policy and inform decisions about HE system and HEIs • Determine national/international competitiveness • Quality, sustainability, relevance and impact of research activity	• System and institutional data re level of intensity, expertise, quality and competence • Performance of HE system and individual institutions • Benchmarking between nationally and worldwide • Indicator of national competitiveness

	• Investor confidence/value-for-money and efficiency • Improve performance and quality • Improve system functionality	• Attraction capacity: recruitment of students, academics and researchers from outside region and internationally • Quality of academic staff and PhD students • Efficiency level: how much output vis-a-vis funding • Research infrastructure: level of use and efficiency
• Ministries of Education/Higher Education or Enterprise and Employment	• Policy and planning • Strategic positioning of HE institutions • Quality, sustainability, relevance and impact of research activity • Research strategy development/management • Investor confidence/value-for-money and efficiency • Quality assurance	• Institutional and discipline/field data re. level of intensity, expertise, quality and competence • Benchmarking against peer institutions, nationally and worldwide • Indicator of national competitiveness • Performance of HE system and individual institutions • Attraction capacity: recruitment of students, academics and researchers from outside region and internationally • Efficiency level: how much output vis-a-vis funding • Research infrastructure: level of use and efficiency
• Local and Regional Governments	• Define local/regional policy and competitiveness • Quality, sustainability, relevance and impact of research activity • Improve integration/collaboration between universities, government and private sector • Improve attraction capacity	• Benchmarking performance and quality of HE system/institutions nationally and worldwide • Indicator of national competitiveness • Attraction capacity: recruitment of students, academics and researchers from outside region and internationally • Efficiency level: how much output vis-a-vis funding
• HE Agencies	• Define policy and inform decisions about HE system and HEIs • Quality, sustainability, relevance and impact of research activity • Determine national/international competitiveness • Investor confidence/value-for-money and efficiency • Improve performance and quality • Improve system functionality	• System and institutional data re level of intensity, expertise, quality and competence • Performance of HE system and individual institutions • Benchmarking between nationally and worldwide • Indicator of national competitiveness • Attraction capacity: recruitment of students, academics and researchers from outside region and internationally • Quality of academic staff and PhD students • Efficiency level: how much output vis-a-vis funding • Research infrastructure: level of use and efficiency
• Other Government Agencies	• Improve and benchmark performance and quality • Aid resource allocation • Investor confidence/value-for-money and efficiency	• Benchmarking performance and quality of HE system institutions nationally and worldwide
ACADEMIC ORGANISATIONS AND ACADEMIES	• Benchmark professional and academic performance and quality	• Academic and discipline/field data re. level of intensity, expertise, quality and competence

	• Student and Academic Recruitment	• Benchmarking against peer institutions, nationally and worldwide • Quality of academic staff and PhD students
INDIVIDUALS		
• Academics and Researchers	• Identify career opportunities • Identify research partners • Identify best research infrastructure and support for research	• Institutional and field data re level of intensity, expertise, quality, competence and sustainability • Performance of individual institution benchmarked against peers in field of interest • Employment conditions • Impact of research on teaching, Staff/student ratio • Institutional research support
• Students	• Inform choice of HEI • Identify career opportunities	• Institutional and field data re level of intensity, expertise, quality, competence and sustainability • Performance of individual institution benchmarked against peers in field of interest • Research capacity of institution and research team, e.g. graduate students/academic ratio, age of PhD students, time to completion, structure/characteristics of PhD programme and support • Graduate career and employment trends • Quality of the research infrastructure • Staff/student ratio
PEER HEIS	• Identify peer HEIs and best research partners	• Institutional and field data re level of intensity, expertise, quality, competence and sustainability • Performance of individual institutions and researchers benchmarked against peers in field of interest • Research capacity of institution and research team • Potential for partnership
INDUSTRY PARTNER ORGANISATIONS		
• Private firms and entrepreneurs	• Quality, sustainability, relevance and impact of research activity • Identify potential partners and expertise • Identify consultancy, technology transfer and knowledge transfer partners and expertise • Identify potential employees	• Institutional and field data re level of intensity, expertise, quality, competence and sustainability • Performance of individual institution benchmarked against peers in field of interest • Competitive positioning of institution and researchers • Trends in graduate employment and competence • Quality of HE programme, and link between research and

		teaching
• Public Organisations	• Quality, sustainability, relevance and impact of research activity • Identify potential partners and expertise • Identify consultancy, technology transfer and knowledge transfer partners and expertise • Identify potential employees	• Institutional and field data re level of intensity, expertise, quality, competence and sustainability • Performance of individual institution benchmarked against peers in field of interest • Competitive positioning of institution and researchers • Trends in graduate employment and competence • Quality of HE programme, and link between research and teaching
• Employers	• Quality, sustainability, relevance and impact of research activity • Identify potential partners and expertise • Identify consultancy, technology transfer and knowledge transfer partners and expertise • Identify potential employees	• Institutional and field data re level of intensity, expertise, quality, competence and sustainability • Performance of individual institution benchmarked against peers in field of interest • Competitive positioning of institution and researchers • Trends in graduate employment and competence • Quality of HE programme, and link between research and teaching
CIVIC SOCIETY AND CIVIC ORGANIZATIONS	• Identify specific expertise and information • Identify potential collaborator • Identify consultancy, technology transfer and knowledge transfer partners	• Institutional and field data re expertise, quality and competence • Peer esteem indicators
MINISTRIES OF HIGHER EDUCATION IN DEVELOPING COUNTRIES	• To help determine which foreign higher education institutions are applicable for overseas scholarships studies. • To help determine research partnerships for knowledge and technology transfer	• Institutional and discipline/field data re. level of intensity, expertise, quality and competence • Competitive positioning of institution and researchers • Trends in graduate employment and competence • Quality of academic staff and PhD students
SPONSORS AND PRIVATE INVESTORS		
• Benefactors/Philanthropists	• Determine institutional performance vis-a-vis national and international competitors • Investor confidence/value-for-money and efficiency • Quality, sustainability, relevance and impact of research activity • Quality of academic staff and PhD student • Contributor to own brand image	• Institutional data re level of quality and international competitiveness • Benchmarking between nationally and worldwide • Quality of academic staff and PhD students
• Alumni	• Determine institutional performance vis-a-vis national and international competitors	• Institutional data re level of quality and international competitiveness

	• Investor confidence/value-for-money and efficiency • Quality, sustainability, relevance and impact of research activity • Quality of academic staff and PhD student • Reflect pride and career aspirations/reputation	• Benchmarking between nationally and worldwide • Quality of academic staff and PhD students
PUBLIC OPINION	• Determine institutional performance vis-a-vis national and international competitors • Quality, sustainability, relevance and impact of research activity • Student choice and career opportunities • Investor/parental confidence and value-for-money	• Institutional data re. level of intensity, expertise, quality and competence • Benchmarking against peer institutions, nationally and worldwide • Indicator of national competitiveness • Performance of HE system and individual institutions • Efficiency level: how much output vis-a-vis funding

3.5 Summary

The Expert Group recognizes that new knowledge is no longer divided strictly between basic and applied activity but includes all research across the spectrum from curiosity-driven to user-led, from blue-sky to practice-based. It involves traditional and new disciplines, and qualitative, quantitative and practice-based methodologies. Research is increasingly conducted through participation in bi-lateral, inter-regional and global networks, involving mono-disciplinary, inter- and multi- and trans- disciplinary forms of inquiry and teams of researchers. Reflecting this complex, iterative and interactive process, new knowledge is disclosed in a wide variety of research outlets and outputs, from peer-reviewed articles to artefacts and prototypes, and including translations, software, encyclopaedia entries, research or technical reports, legal cases and maps.

Drawing on the experience of rankings and existent research assessment exercises, Table 2 presents a comprehensive survey of the wide range of stakeholders and uses to which information on research performance and quality are required and used. Four key points arise from this analysis:

1. Individual user groups have a wide range of AUBR-related requirements.

2. There is overlap between the many requirements of user groups, but some needs are specific to particular groups. For example, governments and government agencies share a requirement for information to aid strategic planning and management, while academics and researchers want information that can aid career choices and research opportunities; universities and government are interested in improving performance and quality while industry and employer groups want to be able to identify potential employees.

3. Some of the required data may be readily available or relatively easy to obtain, while other data are either not available or are available in limited circumstances, which makes comparability across universities or countries difficult. For example, bibliometric data on peer-reviewed publications are available commercially, but there is no similar information available for the wide range of research outputs which would enable cross-national benchmarking; information on trends in graduate employment and competence are not available or counted in a similar manner to facilitate comparability.

4. Existing experience also illustrates that while research performance data may be collected for one purpose, it is often used by other stakeholder groups for very different purposes. This is evident in the way the media often reinterprets or re-tabulates research data as a ‘league table’ or ranking.

These are significant findings, and are testament of the extent to which transparency and accountability of publicly-funded university-based research have become core requirements on higher education. Chapter 4 will discuss some of these issues in more depth.

4 Measuring University-Based Research

This chapter examines the most important characteristics of research assessment. It describes different types of indicators – sometimes referred to as metrics – and the different methods used to measure research activity. It also explains differences in disciplinary practice, and how indicators can positively or negatively affect these differences. Particular attention is given to describing how indicators can be used to measure social and economic impact and benefits. Knowledge clusters are proposed as the optimum unit of assessment. Peer review is also explained. Finally, there is a comprehensive summary of the most commonly used indicators, and their positive and negative features.

4.1 Indicators

Indicators or metrics measure the various components of research activity, including inputs, process, outputs, outcomes and impact and benefits.

•*Input* indicators measure resources, human, physical and financial, devoted to research. Typical examples are the number of (academic) staff employed or revenues such as competitive, project funding for research.

•*Process* indicators measure how research is conducted, including its management and evaluation. A typical example is the total of human resources employed by university departments, offices or affiliated agencies to support and fulfil technology transfer activities.

•*Output* indicators measure the quantity of research products. Typical examples are the number of papers published or the number of PhDs delivered.

•*Outcome* relates to a level of performance, or achievement, for instance the contribution research makes to the advancement of scientific-scholarly knowledge.

•*Impact and benefits* refers to the contribution of research outcomes for society, culture, the environment and/or the economy. See below for further discussion.

The choice of indicators reflects the value judgements and priorities of the promoter. There is no such thing as an objective indicator, because indicators are rarely a direct measurement. Rather they are proxies. For example, citations and publications are a proxy for academic quality; graduate employment is a proxy for the adequacy of graduates for labour-market requirements; budget and research expenditure is a proxy for the quality of the infrastructure; and Nobel or similar awards are a proxy for the quality of research or academic excellence.

Some research assessment exercises (and rankings) assign different weightings or values to the various indicators. In this way, some components of research activity are valued more important than other activity. This may affect the way in which particular disciplines are treated, as discussed below.

The choice of indicators and weightings is therefore vital. They must be fit-for-purpose, appropriate and verifiable. They should be useful to allow decision-making by both internal and external users, and facilitate comparisons over time and across different types of universities. Indicators should be unaffected by any bias arising from the interests of the parties involved in the research assessment exercise. They must also instil trust. In other words, those being assessed need to have confidence that the indicators are appropriate

and truthful. Too few indicators can lead to distortion, while too many can make the exercise too complicated and costly. Ultimately, the process must balance fairness with feasibility, and the results must be replicable, capable of being reproduced in order to be authenticated (European Commission, 2006; Cañibano et al. 2002).

4.2 Indicators and Disciplinary Practice

As discussed in section 3.2 above, different disciplines and research fields have different publication and dissemination practices which can be positively or negatively affected by the choice of indicators.

•**Differences in publication and citation practices**. Publication and citation practices differ significantly from one discipline to another. In some fields, researchers may publish several research articles per year, while in other fields one monograph every 5 years may be appropriate. Citation frequencies also differ across disciplines. This has direct consequences for the journal impact factors published, for example, by Thomson Reuters in its *Journal Citation Reports*. In mathematics, a journal impact factor of 1.0 is high whereas in biochemistry journals with an impact factor of 1.0 is in the lower range. In the social sciences and humanities, journals tend to have impact factors below 1.0.

•***The number of authors per publication***. In medicine, biology and psychology there are usually three to six authors per publication although this could extend to 50 or more for papers on physical acceleration experiment. In contrast, the humanities (history, cultural studies etc.) usually have only one author. The length of a publication has a low correlation with the number of authors: short articles often have several authors, a monograph often only one.

•***Hierarchy of publication outlets***. Some scientific fields (e.g. medicine) rank journals according to their citation impact factors. But in other fields, e.g. social sciences and humanities, there is less concern with the hierarchy of journals and publishers. There is some controversy about the practice of ranking journals, and whether it reflects and/or confirms academic orthodoxy or codifies a field of science.

•***The scope of research***. Some research questions deal with universal phenomena, others have a clear local dimension. While history, ethnography, literary studies, sociology, pedagogy or linguistics address global theoretical questions, they also have strong local or regional dimensions. The scope of research inevitably influences the incentive to publish internationally. This can affect 'national' disciplines, e.g. studies on Portuguese history, literature, language, law, which may not receive fair and equal treatment from the assessors in comparison with 'global disciplines'. An alternative view argues that scientific-scholarly research work, regardless of discipline, should produce universal knowledge and/or explain phenomena or concepts on the basis of general laws or principles – and thus step across a purely local or national viewpoint.

•***The language of a publication***. In the natural, life and technical sciences, English is the dominant language. Certain parts of social sciences and humanities however are more likely to consider issues that have significant or primary national relevance and to publish in the national language.

•***The time span of relevant research***. There are big differences in the time span over which

the research is relevant and cited by other researchers. In some quickly developing areas, research carried out 3-4 years ago tends to be out-of-date and therefore no longer cited; in other areas, studies written five, ten, a hundred or even two thousand years ago are still relevant and can be used and cited. In the natural and life sciences, the time span is normally 5-10 years while in the social sciences and humanities sometimes 10 years is considered too short.

The following topics refer to general issues rather than differences across disciplines.

•***Past performance vs. potential.*** Research assessment practices, especially those which are indicator-based, measure and rely upon past performance both as a method of recognising achievement and as a guide to future performance. 'Old' achievements can be assessed more adequately than 'recent' ones. From a policy point of view, achievements made in the distant past may not be policy relevant. This approach is also unable to capture or identify research potential, new knowledge clusters or fields of investigation.

•***Diversity of Research Missions***. The growing complexity of knowledge and society has corresponded with blurring boundaries between vocational and classical higher education institutions, and between research and development. Simplistic distinctions between basic and applied research have been replaced by greater emphasis on strategic, regional and/or field specialisation. This diversity of research mission is reflected in the wide range of research outputs and outlets mapped across the full spectrum from discovery to knowledge transfer to innovation.

•***Impact of Research on Teaching/Impact of Teaching on Research.*** Many governments and universities strongly support the interconnection between teaching and research as one of the core principles of higher education. A key question for assessment is how to show the validity of the teaching-research nexus. Some assessment processes ask for evidence of new curriculum, changes in pedagogy and new lines of inquiry. Usually this evidence is supported by self-evaluation or case studies, but more work is required.

4.3 Unit of Assessment: Knowledge Clusters

There are many types of research assessment processes at the university and national level, focusing on different institutional or cognitive units. The definition of the unit of assessment depends upon the objectives of the assessment.

There are substantial differences between disciplines regarding what constitutes the natural organizational unit of research or assessment. In the life or physical sciences, the basic unit tends to be the research group. In contrast, in many parts of the humanities, research tends to be conducted on an individual basis, and the individual constitutes the natural unit of research. In clinical medicine, the unit tends to be a multi-disciplinary project group, and one individual can participate in several groups.

There are different ways to classify scientific-scholarly research according to field of science. Table 3 below presents one way to classify 'broad disciplines' according to the ISCED 97 classification made by the UNESCO Institute for Statistics.

Table 3 One Way to Classify Research Activities

ISCED 97 (UNESCO Institute for Statistics)

General Programmes
Education
Humanities and Arts
Social Sciences, Business and Law
Science, Mathematics and Computing
Engineering, Manufacturing and Construction
Agriculture and Veterinary Sciences
Health and Welfare
Services

In addition, universities organise themselves differently for a complex set of reasons, including history, mission, finance, alignment with national or regional priorities, interdisciplinary thematics, etc. Accordingly, it is often difficult to simply compare or benchmark performance on the basis of departments or faculties.

The Expert Group has adopted the concept of 'knowledge cluster' as the basic unit of assessment. A typical knowledge cluster comprises a group of researchers sharing a common field of investigation. Its specific composition is defined as the most appropriate field of study for the specific purpose of the research assessment exercise. In some cases, the knowledge cluster may be related to discipline or field of science, but this categorisation may fail to recognise new or emerging research fields. For example, the knowledge cluster may represent a new or inter-disciplinary issue-driven cluster or an inter-institutional network.

The unit of assessment should always allow for aggregation to the departmental or institutional level as required. This is especially important for international comparability.

4.4 Bibliometric Methods

Bibliometric data is an important method to quantify research activity in terms of that what is published as peer-reviewed journal articles. Commercial bibliometric products are commonly used to capture this information, and it is widely used by university rankings and research assessment.

Thomson Reuters *Web of Science* covers over 9,000 international and regional journals and book series in the natural sciences, social sciences, and arts and humanities. According to its website, 3,000 of these journals accounts for about 75% of published articles and over 90% of cited articles. Journals are selected taking into consideration their basic publishing standards, use of peer review, editorial content, and international diversity of its authorship. S*copus* is a 'flagship product' of Elsevier's Science & Technology Division, with an abstract and citation database of research literature and quality web sources covering almost 18,000 peer-reviewed journals from more than 5,000 publishers. According to its webpage, the database includes extensive conference coverage (3.6 million conference papers), 600 trade publications and 350 book series plus 23 million patent records from 5 patent offices.

The key impediment to bibliometric accounting is that different disciplines produce different types of research outputs which are not all easily recorded. Books, book chapters and

published refereed conference papers are not easily compiled nor recognised as equivalent. Cross-disciplinary and collaborative research is also difficult to categorise. See sections 3.3 and 4.2 above for further discussion of these issues.

4.5 Peer Review

Assessing research quality requires a detailed understanding of the knowledge cluster, in order to evaluate the methodological soundness of the research and the (potential) significance of its contribution to knowledge. Only peers tend to have such an understanding, and this is why peer review has always been an important quality control instrument.

But peer review also has its limitations and biases. Evaluators may be influenced by competitive pressures, including possible implications for their own work or that of their colleagues. They may evaluate research in terms of what they know, and therefore they can act as conservative 'gatekeepers'. In this way, novel and challenging ideas can be marginalised because they challenge established ideas. Finally, peers tend to conform to conventionally accepted patterns of belief, and may, for instance, be influenced by a researcher's reputation rather than his or her actual contribution to knowledge.

4.6 Self-evaluation

Self-assessment is a form of self-reflection which involves critically reviewing the quality of one's own performance and provision. It may be undertaken on an individual basis or, in the context of research assessment, on a collective basis, e.g. involving the knowledge cluster or research team. As a process, it involves the preparation of documentation that reflects on performance over the period in question. This enables the research to be placed within the context of the distinctive research mission and strategy, and encourages consideration of achievements and disappointments.

Self-evaluation benefits from involving researchers in the process of self-knowledge. It encourages them to become involved in helping define 'excellence' and setting the strategy for improvement. In this way, it reinforces internal quality processes.

4.7 Research Ethics

Research should always be conducted in a way which promotes and promulgates good ethical practice, emphasizes integrity and rigour and sustains a culture in which the following guiding principles are understood and observed:

- Honesty, Openness and Fairness
- Confidentiality
- Conflict of Interest
- Respect for Human Subjects
- Respect for Animal Subjects

- Assessment of Risk and Benefits

Research should follow 'good practice' with respect to avoiding plagiarism, accurately documenting results, securely storing primary data, acknowledging the role of collaborators and other participants, and ensuring professional behaviour between supervisor and research students.

Today, completion of an ethical statement or formal ethical approval by a university or national Research Ethics Committee is required by most funding organisations, including the EU. Most research ethics statements follow guidelines established by the *Declaration of Helsinki*, first adopted in 1964 by the World Medical Association, and revised six times, most recently in 2008.

Research assessment processes should ensure that good ethical practice is embedded within the research culture of the knowledge cluster, and the university. This may include consideration of the impact and benefits of the research being conducted.

This issue needs to be further explored, and indicators developed. Peer review would be appropriate.

4.8 Social and Economic Impact and Benefits

Traditionally research assessment has focused on input and output indicators, and measured impact by counting citations as an indication of how knowledge builds upon itself. In response to the wider role and responsibility of university-based research, more attention is being placed on its outcome and benefits, especially its social, economic, cultural and environmental impact.

The purpose of assessing the impact is to gauge the contribution that university-based research makes to society and the economy. This may take of the form of evidence for policymaking, social improvements or the translation of research into cost-effective, practical, policy- and technology-based interventions that improve people's lives. In some instances, this may involve assessing the value, purpose, integrity and ethicality of the research.

It is also important to inform government, industry, business and the community about the results of public investment in research. This arises from the need to assure stakeholders that publicly-funded university research is valuable and has been rigorously assessed through internationally recognised processes.

Measuring impact and benefits is an emerging methodology, and additional work needs to be done in order to identify appropriate indicators, but also develop mechanisms to collect accurate and comparable data. The indicators can be quantitative and qualitative. The latter may involve end-user reviews which are similar to the role that peer-review performs when assessing traditional academic outputs, in addition to self-evaluation reports.

•***Economic Benefits***, e.g. improved productivity; adding to economic growth and wealth creation; enhancing the skills base; increased employment; reduced costs; increased innovation capability and global competitiveness; improvements in service delivery; as well as unquantifiable economic returns resulting from social and public policy adjustments.

•***Social Benefits***, e.g. improving people's health and quality of life; stimulating new approaches to social issues; changes in community attitudes; influence upon developments or questions in society at large; informed public debate and improved policy-making; enhancing the knowledge and understanding of the nation; improved equity; improvements in health, safety and security; improved social attachment; and improvements in the level and security of political rights.

•***Environmental Benefits***, e.g**.** improvements in environment and lifestyle; reduced waste and pollution; improved management of natural resources; reduced consumption of fossil fuels; uptake of recycling techniques; reduced environmental risk; preservation initiatives; conservation of biodiversity; enhancement of ecosystem services; improved plant and animal varieties; and adaptation to climate change.

•***Cultural Benefits****, e.g.* supporting greater understanding of where we have come from, and who and what we are as a nation and society; understanding how we relate to other societies and cultures; stimulating creativity within the community; contributing to cultural preservation and enrichment; and bringing new ideas and new modes of experience to the nation.

Further details are available in Appendix III.

4.9 Indicators and Their Dimensions

Indicators – or metrics – are used to measure different aspects or dimensions of research activity. Dimensions go beyond scientific quality to capture relevance, impact, resources and infrastructure. Table 4 below describes the most commonly used indicators. Each description is followed by a short commentary identifying some positive (pro/potentialities) and negative (con/limitations) features. The last column suggests some steps which are required to make the indicator more robust.

The indicators are categorised according to what they aim to measure:

•Research Productivity;

•Quality and Scholarly Impact;

•Innovation and Social Benefits;

•Sustainability and Scale;

•Research Infrastructure.

Table 4 should be read in conjunction with Appendix III.

Table 4 Overview of Indicators, and Some Positive (pro/potentialities) and Negative (con/limitations) Features

INDICATORS	DESCRIPTION	PRO/POTENTIALITIES	CON/LIMITATIONS	WHAT DEVELOPMENT IS REQUIRED
RESEARCH PRODUCTIVITY				
Research publications and outputs	A count of publications and other research outputs.	Depending on purpose only selected types of publications can be counted. Publishing is vital for progress in science scholarship.	Different disciplines produce different types of research outputs. Emphasis on quantity of publication.	Suitable data bases for a variety of disciplines and research related outputs, especially in social sciences and humanities.
Research outputs per 'Research Academic' staff	Number of publications and other research outputs per academic staff or full-time equivalent (FTE).	Supports cross-institutional comparisons, adjusted for scale of institution.	Comparable definition of 'Academic Staff' and 'Research Time' can be difficult.	Agreement on definition of 'Research Academic'.
QUALITY AND SCHOLARLY IMPACT				
Number and percentage of publications in top-ranked, high impact journals	The number or percentage of journal articles published in the top-ranked, high impact journals for the fields of research.	In the exact sciences, peers tend to consider citation impact a relevant aspect in assessments of research performance. Widely used, especially in the exact sciences which tend to be well covered. Data must be accurate and verified.	Although one of the most popular indicators, it is not always the most appropriate one. Especially in social sciences and humanities, expert rankings do not correlate very well with impact factors. In these fields and in engineering, other sources are important as well (books, proceedings).	Discipline specific journal rankings, especially in social sciences and humanities, based on expert opinion in combination with indicators. Value of developing a ranking or hierarchy of scientific-scholarly publications.
Citations	Citation data are derived from citation indexes, i.e. databases that do not only contain meta data on included	In the exact sciences, peers tend to consider citation impact a relevant aspect in	Citations reflect intellectual influence but do not fully coincide with research quality.	Expansion of existing databases and creation of new databases (e.g. based on data from institutional repositories) will

INDICATORS	DESCRIPTION	PRO/POTENTIALITIES	CON/LIMITATIONS	WHAT DEVELOPMENT IS REQUIRED
	publications but also their reference lists. Principal indexes are Web of Science, Scopus and Google Scholar.	assessments of research performance. Widely used, especially in the exact sciences which tend to be well covered, although the most popular indicators are not always the most appropriate ones. Data must be accurate and verified.	Are of limited value in disciplines not well covered by the citation indexes, especially certain parts of social sciences, humanities and engineering.	improve the value of this indicator and coverage of disciplines. Theoretical research into the meaning of citations (clusters) in social sciences and humanities.
Number Keynote Addresses at Nat'l/Int'l Conferences	A count of the number of invited and keynote addresses given at national and international conferences	Used as proxy for quality, impact and peer-esteem. Data can be verifiable by conference programme.	No agreed equivalences that apply internationally and facilitate comparison across disciplines.	This will probably require direct entry by researchers. A list of internationally comparable items for different disciplines might help a lot.
Number Prestigious Nat'l/int'l Awards and Prizes	A count of the number of prestigious national and international prizes won either in total or per academic staff.	Used as an indicator of research quality and impact. Data is verifiable.	No agreed equivalences that apply internationally and facilitate comparison across disciplines.	Unless lists are publically available this will require direct entry by researchers. A list of internationally comparable items for different disciplines might help a lot.
International Visiting Research Appointments	A count of the number of visiting appointment at other academic and/or non-academic agencies and organisations.	Visiting Appointments provide indication of peer esteem or support by the academic community. Numbers are verifiable.	No agreed equivalences that apply internationally and facilitate comparison across disciplines.	Will probably require direct entry by researchers.
Editorial and Refereeing for Prestigious National/International journals/publishers	A count of the number of national and international appointments as editor, member of editorial board or as reviewer	An indicator of the extent to which the researcher's opinion is highly regarded by the academic community. Data is verifiable	No agreed equivalences that apply internationally and facilitate comparison across disciplines.	Unless lists are publically available this will require direct entry by researchers. A list of internationally comparable items for different disciplines might help a lot.

INDICATORS	DESCRIPTION	PRO/POTENTIALITIES	CON/LIMITATIONS	WHAT DEVELOPMENT IS REQUIRED
INNOVATION AND SOCIAL BENEFIT				
External research income .	Level of funding attracted by researchers and universities from external sources, including competitive grants and research income from government, industry, business and community organisations.	Comparable data, verifiable through audit, is useful for comparing research performance across the system and within universities. Willingness of industry to pay for research is a useful indicator of its anticipated contribution to innovation and the economy.	Levels of external funding vary greatly across disciplines. For example, in countries where over half the total pool of funding is allocated to medical research, universities that do not have Medical Faculties will inevitably secure less funding than those with Medical Faculties. Data collection may be difficult in case of funding by end users because this information is not known to the University administration.	Agree international comparative data base.
Number and percentage competitive grants won	Level of funding won competitively – this is a sub-set of the indicator above.	Comparable data, verifiable through audit, is useful for comparing research performance across the system and within universities.	Levels of external funding vary greatly across disciplines. See above.	Agree international comparative data base.
Research income per academic staff or FTE	Research income per academic staff or FTE supports cross-institutional comparisons, adjusted for scale of institution.	Important measure of research activity.	Comparability is dependent upon institutional mission, context and discipline.	Data needs to be adjusted to scale and mission of university.
Employability of PhD graduates	Industry employment of PhD graduates can be an indicator of the contribution of research to the highly	Used to measure the quality of the graduates, and impact of research on	Employability can be sensitive to other factors, such as the regional or national economy.	Important to develop methods to track graduate employability and career paths. Harmonise the stage(s) post-graduation at

INDICATORS	DESCRIPTION	PRO/POTENTIALITIES	CON/LIMITATIONS	WHAT DEVELOPMENT IS REQUIRED
	educated & skilled workforce.	teaching.	Career paths and opportunities can differ for different disciplines.	which data is collected.
Commercialisation of research-generated intellectual property (IP)	Provides measure of the extent of income from commercialisation of intellectual property created through patents, licences or start ups.	This is an area of increasing significance to policy makers. Indicator is an important link between IP, commercialisation and economic benefits.	Patents are a very poor indicator of commercialisation. They are sensitive to national context – and to discipline.	Databank on university related inventions should be developed.
End-user Esteem	Includes policy, technical or commissioned reports; consultancy and external contracts; architectural or design awards; etc.	Willingness of external stakeholders to use and/or pay for research is a useful indicator of its anticipated contribution to innovation and the economy.	Different opportunities for different disciplines. Lack of agreed basis of capturing data and comparability could undermine legitimacy.	Agree basis of international comparability and verifiability.
Number and percentage funding from End-users (e.g. industry, professions, government, community)	Provides measure of the extent of income from external-commissioned or contracted work.	This is an area of increasing significance to policy makers. Indicator is an important link between research and social and economic benefits.	Different opportunities for different disciplines. Lack of agreed basis of capturing data and comparability could undermine legitimacy.	Agree basis of international comparability and verifiability.
SUSTAINABILITY AND SCALE				
Postgraduate Research Student Load	The ratio of research students (or PhD students) per academic staff or per 'Research Active' staff.	Key indicator of research intensity, indicating the scale of the research enterprise.	Practices differ across disciplines – large research teams are a common feature of the bio- and medical sciences.	Agree basis of international comparability and verifiability.
Involvement of early career	Number or percentage of early stage researchers involved in research	An indicator of research intensity, the scale of the	Practices differ across disciplines – large research	Agree definition of 'early career researcher', and basis of international

INDICATORS	DESCRIPTION	PRO/POTENTIALITIES	CON/LIMITATIONS	WHAT DEVELOPMENT IS REQUIRED
researchers in teams	activity.	research enterprise, and future activity.	teams are a common feature of the bio- and medical sciences.	comparability and verifiability.
Number of Collaborations and Partnerships	A count of national and international collaboration with other universities and/or with public-private and NGOs, etc.	Because research is increasingly conducted in collaborative teams, nationally and internationally, this is an important indicator of research involvement and scale of activity.	Can be difficult to capture and verify the data due to lack of clarity as to what is being measured.	Agree precise definition, inter alia: university-university, university-external stakeholder, national, European or international.
Doctoral Completions	The number PhD and equivalent research doctorates and, as appropriate, research Masters degree completions.	Data is verifiable by universities although there can be a time lag.	Rates of completion may differ across disciplines. Different disciplines may prioritise masters and PhD activity.	Require common standard for doctorates
RESEARCH INFRASTRUCTURE				
Research active academics	Number or equivalent full-time (FTE) of 'research active' academics employed by a university. 'Research active' is established by setting threshold levels of performance for a specific period.	Important indicator of research capability.	No clear definition of 'Research Active'.	Common definition and international comparability of 'Research Active'.
Percentage 'Research Active' per total academic staff	Ratio of the number of 'Research Active' per total academic staff.	Indicator of research intensity.	No clear definition of 'Research Active'.	Common definition and international comparability of 'Research Active'.
Total R&D investment	Total investment in university-based R & D (research and development) from all sources, including external research income and university resourcing of research, including	Investment in research is a strong predictor of research performance.	Difficult to get valid, comparable institutional data, even within the same institution.	Agree basis on which to calculate full cost of research investment.

INDICATORS	DESCRIPTION	PRO/POTENTIALITIES	CON/LIMITATIONS	WHAT DEVELOPMENT IS REQUIRED
	salaries and overheads.		Can be difficult to fully calculate university resourcing of research.	
Research Infrastructure and Facilities	Number of research laboratories, Books in the library and/or electronic journal access, super-computing access, etc.	Information provided at the level of the institution.	Difficult to get valid, comparable data. Favours older, well-endowed universities.	Develop appropriate comparative indicators.
Research Ethics	Comprehensive process ensuring good ethical practice is promoted and promulgated.	Important measure of research rigour and integrity, and the effect and purpose of the research. Peer Review would be most useful.	Ethical statements regarding the use of research and the source of research funding, e.g. tobacco or armaments, can be very controversial.	Develop appropriate indicators to ensure good ethical practice is promoted without interfering in processes of discovery.

4.10 Summary

The Expert Group notes that indicators are chosen as a method of measuring the various aspects of the research process. They are rarely a direct measurement but rather are proxies. Indicators should be fit-for-purpose and verifiable; they must be fair, appropriate and capable of facilitating comparisons across disciplines and institutions. All indicators have positive and negative characteristics. It is therefore necessary to consider the merit and fit-for-purpose of each indicator.

Because of differences in disciplinary research practice, great care should be taken in the choice, use and interpretation of indicators. Research assessment should take account of the social, economic, environmental and cultural context in which the university operates and the research is conducted. In addition, assessment of its impact and benefits should be included in recognition of the wider role and responsibility of university-based research.

Finally, knowledge clusters should be the basic unit of assessment, for five main reasons:

1. Universities differ in their internal structures;

2. Different disciplines tend to have different types of research units;

3. It enables new and emerging disciplines, and interdisciplinary research clusters to be assessed;

4. The choice knowledge cluster is flexible enough to be adapted to the purpose of the assessment exercise; and

5. Knowledge clusters facilitate aggregation to the department or institutional level as required.

Chapter 4 has shown that indicators are complex instruments for measuring research performance and quality. No single set of indicators is capable of capturing this complexity, and therefore research assessment based upon single indicators is likely to lead to distortion. Chapter 5 will propose a methodology and tool for making sense out of this complexity.

5 A Proposed Framework for Research Assessment

This chapter presents a new approach to research assessment. It first illustrates some characteristics of 'good practice' which are found in the case studies of existent research assessment practice, in Appendix IV. Drawing upon that experience plus the basic principles outlined in Chapter 3, and the characteristics of research assessment described in Chapter 4, a *Multidimensional Research Assessment Matrix* is proposed.

The Expert Group recommends that this Matrix be used to form the basis of a full-fledged personalised and web-enabled Toolkit. Examples are provided below in section 5.3 illustrating how it could be implemented by the various user groups identified above in section 3.4 and the accompanying Table 2.

5.1 Lessons from Existing Practice

Many lessons about research assessment can be learned from examining ongoing practice. Appendix IV provides a series of case studies which draws on the expertise of the members of the Expert Group. This section highlights some key aspects of that experience which should be embedded in any 'good practice' model; section 6.1 refers to aspects of current practice that have brought about unintended consequences.

•***Consultation with Researchers and Universities:*** Development of both the Research Quality Framework (RQF) and the Excellence in Research for Australia (ERA) involves extensive consultation to ensure procedural fairness, transparency and acceptance by key stakeholders.

•***Data Collection through Digital Repositories*****:** Technology provides an easy way to store and access research for the actual research assessment process, as well as ensuring wider dissemination. Although digital institutional repositories (Australia) and web-based tools (e.g. Webometrics and Google Scholar) currently cover only a limited part of university-based research outputs, in the future they could become important sources of information and overcome some of the limitations of traditional bibliometric databases.

Digital repositories and web-based tools can facilitate scientific collaboration in line with the movement for open science. They promote transparency in experimental methodology, observation, and collection of data; public availability and reusability of scientific data; and public accessibility and transparency of scientific communication.

•***Peer Review Panels:*** Several case studies underscore the importance of peer review panels. The process helps ensure a broader understanding of the research and its contribution to knowledge, including the importance of new disciplines and inter-disciplinarity. In Finland, France, the Netherlands and the UK, panels include international experts; in Finland, France and the Netherlands, this also involves visits to the university. Peer review panels are also used to assess the quality of research outputs and outlets of individual researchers for career promotion, such as in Spain.

•***Indicators:*** All systems use bibliometric indicators, although many balance this with other information about the research environment, research strategy and management, and impact on teaching. The Netherlands combines retrospective and prospective analysis. The

UK and Australia are adopting an indicator-based system with lighter-touch peer review and social impact.

•***Purpose:*** All research assessment exercises aim to improve research performance and quality. The UK system is directly tied to resource allocation, while the Netherlands system is used primarily to improve quality. The French system represents an interesting mixture of the two.

•***Self-evaluation:*** Aalto and Helsinki Universities, Finland, and the Netherlands include self-evaluation as a key component in the process.

•***Social and Economic Impact and Benefits:*** Several countries and universities are experimenting with measuring societal impact, demonstrated through case studies, end-user opinion, and appropriate indicators. This is most notable in the Australian Research Quality Framework (RQF), developed in 2005-07, Aalto University in Finland, and the Netherlands.

•***Unit of Assessment:*** Many case studies focus on the research discipline (Australia) or research unit (Germany, Forschungsrating), making it possible to illustrate differences in research quality within individual universities. The CHE ranking in Germany deliberately does not enable aggregation across a whole university.

Although the case studies represent very different systems and objectives, they share some common positive attributes. Research assessment is seen as aiding strategic planning, international benchmarking, and bringing about greater cohesion and organisation among discipline groupings.

Drawing on the experience of existent research assessment exercises and of rankings, the Expert Group has developed a *Multidimensional Research Assessment Matrix* – discussed below.

5.2 Framework for Research Assessment

Assessing university-based research is a complex process. It is most often undertaken to, inter alia, improve research performance and quality, allocate resources, drive research mission differentiation or promote innovation and community engagement.

No single set of indicators can meet all these requirements or provide all solutions. Each purpose requires different data. Some requirements demand extremely detailed and robust data on research outputs; other requirements demand only a few, relatively simple indicators. All indicators have advantages and disadvantages, and there are limitations to all assessment exercises (see Chapter 6).

Indicators designed to meet a particular objective or inform one target group may not be adequate for other purposes or target groups. Diverse institutional missions, and different policy environments and objectives require different assessment processes and indicators. Likewise, the range of people and organizations requiring information about university-based research is growing. Each group has specific but also overlapping requirements.

The optimum method of research assessment is to combine qualitative and quantitative indicators, e.g. indicator-based data with peer or end-user review. There are several advantages to this approach:

•Indicators provide peer experts with condensed, systematic, verified and 'objective' information on the research performance of the knowledge cluster.

•The results provided by indicators are never self-evident but must be evaluated and interpreted in the context of the discipline, national circumstances, the university, etc. This can only be done by experts.

•Indicators aid good judgement by supporting or challenging peer assumptions, thereby making the assessment process more transparent.

•Indicators can provide additional, objective information on research performance from a global perspective.

5.3 Multidimensional Research Assessment Matrix

To fairly and accurately assess the full range of research activity requires a multidimensional approach which reflects different uses and users. The *Multidimensional Research Assessment Matrix* is a methodology which can help make sense out of this complexity. Table 5 presents the Matrix in summary format, illustrating how particular sets of indicators can be used to meet different policy needs. The actual choice of indicators depends upon the purpose and the availability of data.

There are four important steps:

Step 1: Define the purpose and audience of the research assessment exercise;

Step 2: Involve the universities and other users, as appropriate, in Step 1 above;

Step 3: Identify the appropriate indicators;

Step 4: Identify the range of actions and decisions to be taken upon completion of assessment.

Table 5

Multidimensional Research Assessment Matrix

PURPOSE	RESEARCH PRODUCTIVITY	QUALITY AND SCHOLARLY IMPACT	INNOVATION AND SOCIAL BENEFIT	SUSTAINABILITY AND SCALE	RESEARCH INFRASTRUCTURE
Allocate Resources	• Research output/bibliometric data	• Citation data • Peer review • Keynote, awards, etc.	• Research income		• 'research active' as percentage of total academic staff; libraries, equipment, etc.
Drive Research Mission Differentiation	• Research output/bibliometric data • Output per research academic	• Peer review • Self-evaluation	• Ratio of research income: teaching income • External research income	• Ratio of undergraduate: master/phd students	
Increase Regional/ Community Engagement	• Publications, Policy Reports, etc.	• End User Reviews • Keynote, Media Awards, etc.	• Percentage Funding from End-users • Patents, Licenses, Spin-Offs	• Number of collaborations and partnerships	
Improve Research Performance	• Research output/bibliometric data	• Citation data • Number and percentage publication in top-ranked, high impact journals • Peer review			
Assess Value-For-Money or Cost-Benefit of Research	• Research output/bibliometric data • Output per research academic	• Peer review and/or citation data • Commercialisation data • End user reviews	• Social, economic, cultural and environmental impact/benefits indicators • External research income • Employability of PhD graduates	• Number of collaborations and partnerships	
Encourage International Co-operation	• Research Output/Bibliometric data with focus on European & International collaborations		• Percentage of Research Income from International Sources	• Number of collaborations and partnerships	
Increase Multi-disciplinary Research	• Research Output/Bibliometric data with focus on interdisciplinary fields	• Peer Review • Self-Evaluation	• New research fields, interdisciplinary teaching programmes, etc.		• Research Conducted by People from Different Disciplines

Some illustrative scenarios follow.

IF you want to use research assessment to ALLOCATE RESOURCES, then what is required is:

- Regular cadastral census of research outputs
- Peer review to determine quality and/or citation data to determine impact
- Some measure of research infrastructure / environment – e.g. Libraries, equipment, postgraduate student numbers, etc,
- 'Esteem' factors – e.g. prizes, research income etc.

IF you want to use research assessment to DRIVE RESEARCH MISSION DIFFERENTIATION, then what is required is:

- Data on Research Outputs, including output per academic staff
- Data on ratio of research income: teaching income
- Data on ratio of undergraduate students: master & doctorate research students
- Peer Review Panels
- Self-Evaluation Reports

IF you want to use research assessment to INCREASE REGIONAL/COMMUNITY ENGAGEMENT, then what is required is:

- Data on cooperation agreements with local governments and organisations of the region;
- Data on agreements with other public or private institutions located in the targeted area of influence;
- Indicators of results (publications, policy reports, patents, spin-offs, …..) coming from these agreements;
- Ratio of business or other external funding of research: general research funding;
- Data on 'merit' of research as assessed by end users, rather than peer review;
- Peer Esteem, e.g. expert opinion, professional memberships, media visibility.

IF you want to use research assessment to IMPROVE RESEARCH PERFORMANCE, then what is required is:

- Option A: Use a highly selective resource allocation model dependant on quality of outputs:
    - Regular cadastral census of research outputs
    - Peer review to determine quality and/or Citation data to determine impact
- Option B: Use holistic peer review assessment panels to benchmark performance against international comparators, assisted by simple output indicators.

IF you want to ASSESS VALUE-FOR-MONEY OR THE COST-BENEFIT OF RESEARCH, then what is required is:

- Data on research outputs, including output per academic staff;
- Peer review and/or citation data to determine scholarly impact;

• Indicators of commercialisation of IP;
• Indicators of social, economic, cultural and environmental impact and benefits;
• Ratio of business or other external funding of research: general research funding;
• Indicators of end-user esteem, e.g. policy and commissioned reports, expert opinion, professional memberships, media visibility;
• Data on employability of PhD graduates;
• Data on collaborations and partnerships.

IF you want to use research assessment to ENCOURAGE INTERNATIONAL COOPERATION, then what is required is:

•Data on European and international cooperation agreements;
•Data on joint publications with scholars from other countries;
•Proportion of research funding, domain by domain, coming from overseas research institutions.

If you want to use research assessment TO INCREASE MULTIDISCIPLINARY RESEARCH, then what is required is:

•Use knowledge clusters as unit of assessment by peer review panels;
•Data on output according to knowledge cluster – perhaps using bibliometrics with focus on authors from different disciplines;
•Data on other results, e.g. new research areas, courses or teaching programmes, …) designed together by people from difference disciplines (or Schools, or Faculties…);
•Peer review;
•Self evaluation.

6 Conclusion

6.1 Limitations and Unintended Consequences

University-based research has become one of the most critical factors shaping national competitiveness and university reputation. This situation is likely to intensify as global competition increases further, (public) funding for research is reduced, and human capital formation becomes more vital for innovation and economic growth. Research assessment exercises are likely to become an increasingly important determinant of research quality and status. These factors have contributed to the growing importance now attached to research assessment.

Some governments, public agencies and universities are drawing immediate and direct links between research assessment and resource allocation or accreditation. These and other actions can be viewed positively as part of the broader modernisation agenda, setting priorities and improving performance and public accountability, while others can be viewed as counter-productive. Thus, the choice of indicators, and their purpose, can aid decision-making or lead to distortion.

Unintended consequences can occur when the results are taken in isolation and simple correlations made between cause and effect. Drawing conclusions, for example, about economic performance based upon research assessment can lead to over-concentrating on research – for example, to the detriment of teaching or favouring particular disciplines, especially the sciences. Such consequences are evident in a number of the case studies appended, e.g. France, Norway and Sweden. They illustrate the risk of aligning higher education priorities and resources to match indicators.

Simple ranking of universities on the basis of bibliometrics or citations can ignore differences between disciplines and between university missions, resources and context. Likewise, the absence of appropriate, verifiable and trustworthy data can undermine the usefulness of cross-national comparisons and benchmarking. Research has found that the results of the Shanghai Jiao Tong *Academic Ranking of World Universities* (ARWU) are not replicable, thus calling into question the comparability and methodology used. The Spanish and Italian examples raise other concerns about the lack of transparency; the names of experts being kept secret (Italy) and uncertainty about the way candidates are evaluated (Spain).

Bibliometric and citation data is by definition backward looking; in other words, it assesses past performance as a proxy for future performance. This approach, while often valid, can easily ignore new and emerging disciplines, young researchers, and new universities. In other words, these indicators cannot easily measure potential. This underlines the need to combine indicators and expert knowledge, and illustrates why a multidimensional system of assessing scientific-scholarly research should balance past performance criteria and the quality of research proposals and future plans.

At the same time, decisions made on the basis of research assessment exercises can also be complex. While it might be appropriate to allocate resources to researchers or universities which have performed best, the alternative could also be appropriate, in other words, to allocate resources to weaker universities in order to build up their capacity. Or, rather than

using resource assessment to drive differentiation, it might be more useful to encourage university self-profiling.

Finally, in recent years, concern has been expressed about the financial and human costs associated with research assessment. This has led to calls for the adoption of simple indicator-based systems. However, the absence of verifiable and accessible cross-national data and common definitions raises questions as to the efficacy of this approach on an international basis given all the limitations that have been identified throughout this report. There have also been reports of 'gaming' or 'intelligent' presentation of data by universities and researchers anxious to ensure a good report.

Because of unintended consequences, the choice of indicators, methodology and data sources are critical.

- Qualitative indicators can easily ignore differences between disciplines;
- Peer reviewers can also act as 'gate-keepers';
- Evaluation systems may encourage behaviour which is contrary to particular policy needs;
- Indicators measure past performance rather than potential;
- Bibliometric methodologies focusing on publications in peer reviewed scientific-scholarly journals fail to capture research activity across the full research innovation eco-system;
- Reliance on data that is easily measured can distort research towards that which is more predictable;
- Emphasis on global impact can undermine the importance of regionally relevant outcomes.

6.2 Good Practice

There is increasing interest across the European Union and worldwide about the contribution and impact that university-based research is having on society and the economy. The rise of worldwide ranking of universities is testament to this interest. This is the background against which the Expert Group on Assessment of University-based Research was established by DG Research in 2008. In the intervening months, the global financial crisis has focused the attention of EU and national governments on the necessity to develop highly-skilled internationally competitive and sustainable smart economies and societies.

In order to be effective, it is vital to balance the objectives of improving performance and quality; driving-up research excellence; providing better and more transparent information to students, potential students and the public; providing the basis for evidence-based policy-making; and assuring confidence for the public and the taxpayer. Because the results of research assessment can carry great significance for university, researcher and student reputation and status, the process must be transparent and the results must be replicable in order to be authenticated.

In line with international practice to ensure wider dissemination and take-up of research into society and the economy, support for 'experiments' with open science, open source and open repositories is vital. Although the added value of these approaches – including

their implications for scientific-scholarly practice and their economic aspects – need to be carefully studied, they could eventually also help overcome some of the limitations inherent in currently available bibliometric and citation databases.

'Good practice' suggests that research assessment should

1. ***Combine indicator-based quantitative data with qualitative information***, for example information based on expert peer assessment. This enables the quantitative information to be tested and validated within the context and purpose of the assessment, with appropriate reference to the discipline and disciplinary practice.

2. ***Recognise important differences across research disciplines***. Peer-reviewed journal articles are the primary publication channel for practically all disciplines, but the complexity of knowledge has led to a diverse range of output formats and outlets.

3. ***Include assessment of impact and benefits.*** Because research does not exist in isolation, assessment should include indicators which are capable of capturing and recognising this. This differs for different disciplines. Stakeholder esteem indicators can show how research is viewed by the wider community.

4. ***Integrate self-evaluation*** as a useful way to include the research community pro-actively in assessing their own contribution, but also as a means of placing the research process – which includes the organization, management, and developments over time – into context and related to institutional mission.

The Expert Group on Assessment of University-based Research has proposed a *Multidimensional Research Assessment Matrix*. Adapting the Matrix to web-based technologies would enable different users to personalise the various dimensions and characteristics to meet their particular policy objective. This would substantially enhance its user-friendliness.

6.3 Contribution to Future Research Assessment Exercises

The AUBR Expert Group hopes that this report will raise awareness of the principles that need to be observed in assessment of university-based research and of the pitfalls that should be avoided. In recognising that research assessment must always be fit-for-purpose, the Expert Group

- Has identified a wide range of users who seek and require research assessment information for a variety of different and often conflicting uses;
- Has recognised that all disciplines have an important contribution to make to the knowledge triangle;
- Holds the view that all indicators have their potentialities and limitations, and research assessment exercises must be cognisant of these from the outset;
- Believes that the *Multidimensional Research Assessment Matrix* provides the optimum basis for strategic decision-making by government and government agencies, universities and other stakeholder organisations;
- Urges governments to invest the requisite resources to ensure the full breadth and depth of university-based research can be fairly and appropriately assessed and benchmarked;

•Encourages a thorough assessment on a regular basis of university-based research along the lines proposed in this report rather than being guided merely by global rankings.

In sum, the Expert Group hopes that this report will serve as a guide to

•Users of information on the quality of university-based research, enabling them to form a judgment as to the adequacy or otherwise of existing and future assessment exercises and systems;

•Decision-makers in higher education, encouraging them to take the assessment of research seriously, but at the same time keeping an eye on the other functions of their institution – teaching and learning, and community engagement;

•Specialists engaged in assessment of university-based research, presenting them with a number of principles that need to inform assessment of university-based research.

The Expert Group also hopes that this report will provide inspiration to the European Commission and Member State governments to launch initiatives and projects designed to generate much needed comparable data and more appropriate and robust scenarios for the assessment of university-based research.

7 Appendix I. Activities and Membership of Expert Group on Assessment of University-Based Research

The members of the Expert Group were selected on the basis of their experience and knowledge of research assessment and higher education, with an emphasis on different institutional and national/regional viewpoints, and a good mixture of academic, industrial and policymaking backgrounds and professional experiences. This group includes fifteen (15) members (including the Chairperson and 2 Rapporteurs) to provide a variety of views and approaches. Thirteen (13) members are from European countries, including a representative of the European Universities Association (EUA), plus two (2) international experts.

MACKIEWCZ Wolfgang (Chairperson)
HAZELKORN Ellen (Rapporteur)
BERGHOFF, Sonja (Rapporteur)
BONACCORSI, Andrea
BORRELL-DAMIAN, Lidia
EMPLIT, Philippe
INZELT, Annamaria
MARKLUND, Goran
MOED, Henk
MUSTAJOKI, Arto
NEWBY, Howard
PEYRAUBE, Alain
ROWLEY, Sue
SALMI, Jamil
SANCHEZ, Paloma

The Expert Group was coordinated by **Adeline Kroll** (Scientific Officer, EC DG/RTD Directorate, *Unit C4 – Universities and Researchers; European Research Area: Knowledge-based economy*) with secretarial assistance from **Celina Pastor Rubio and Alina Hossu**.

Curriculum Vitae of Members of the Expert Group

MACKIEWCZ Wolfgang (Chairperson). **Wolfgang** is director of the Language Centre and Honorary Professor of English Philology at the Freie Universität Berlin (FUB). He is the president of the Conseil Européen pour les Langues/European Language Council (CEL/ELC). He was chair of the SIGMA Scientific Committee on Languages (1994-5), and has coordinated a total of eight EU development and network projects, among them five thematic and dissemination network projects in the area of languages (Socrates-Erasmus Programme; 1996-2007). Over the past 13 years, he has been advisor to both the European Commission and the Council of Europe in a number of capacities. He chaired the Expert Group on the Humanities in FP7, he is the chair of the Expert Advisory Group FP7 Theme 8 and of the Assessment of University-Based Research Expert Group, and he was *rapporteur* of the European Commission's High Level Group on Multilingualism. Wolfgang Mackiewicz studied English and German at FUB and at the University of Leeds, and wrote his PhD thesis on Daniel Defoe's 'Robinson Crusoe'. His research focuses on European language policy and language education policy. He has published numerous articles and chapters in books. He holds three honorary doctorates. Email: erasmspr@zedat.fu-berlin.de.

HAZELKORN Ellen (Rapporteur). **Professor Hazelkorn** is the Director of Research and Enterprise, and Dean of the Graduate Research School, Dublin Institute of Technology, Ireland; she also leads the Higher Education Policy Research Unit. She is a Consultant to the OECD Programme on Institutional Management of Higher Education (IMHE), and is also associated with the International Association of Universities (IAU). Ellen is Rapporteur for the EU Expert Group on Assessment of University-based Research, and a member of National Digital Research Centre (NDRC) Management Board, the Arts, Humanities and Social Sciences Foresight Working Group [Ireland], and the International Advisory Council of the Irish Research Council for the Humanities and Social Sciences. She is a member of the Executive Committee of the Dean and European Academic Network (DEAN), and the Editorial Boards of Higher Education Management and Policy (OECD) and Higher Education Policy (IAU). Professor Hazelkorn has published articles and books on Irish politics and society; digital technologies, gender, work practices and the cultural industries; relations between the media and the state; and higher education policy. Her study, Developing Research in New Institutions, was published by OECD (September, 2005). She is studying the Impact and Influence of Rankings on Higher Education Decision-Making and Academic Behaviour in association with IMHE and IAU [http://www.oecd.org/edu/imhe/rankings], and working with the Institute for Higher Education Policy (USA) on a 'New Agenda for College and University Ranking'. Her book, *Rankings and the Reshaping Higher Education: The Battle for World-Class Excellence*, will be published by Palgrave in 2010. E-mail: Ellen.Hazelkorn@dit.ie

BERGHOFF, Sonja (Rapporteur) **Sonja** is a project manager at the German CHE Centre for Higher Education Development. She is responsible for the research indicators in the different CHE rankings and especially leader of the CHE ResearchRanking for German universities. In 2000 Sonja Berghoff joined the CHE ranking team. She was responsible for the students' survey till 2006 and from 2002 leader of the CHE ResearchRanking of German universities. This project aims at discovering departments or faculties in German universities that are most active in research. Sonja Berghoff was responsible for the bibliometric analyses in all disciplines as well as the aggregation and presentation of the results. Sonja Berghoff studied Statistics at the University of Dortmund and wrote her PhD-Thesis on "Heteroscedasticity and autocorrelation consistent covariance matrix estimation within the linear regression model". She worked as manager for the collaborative research centre 475 " Reduction of complexity in multivariable data structures" for about three and a half years and in this way gained a lot of experience concerning the organisation of third party funding at German universities as well as evaluational aspects during the phase of (re-)application und examination. E-mail: Sonja.Berghoff@CHE-Ranking.de

BONACCORSI, Andrea. **Andrea** is Professor of Economics and Management at the University of Pisa. He is author of papers in the most important journals in economics and policy of science, technology and innovation, and the editor (with C.Daraio) of *Universities and strategic knowledge creation* (Edward Elgar, 2007). He has pioneered the construction of datasets on universities in Europe, using microdata to carry out evidence-based policy analysis. He is member of the HLEG on Assessment of university based research and of the HLEG on the Future of Commission research policy at DG Research, and advisor to the Italian Ministry of Economic Development on Structural Funds for innovation, as well as of several regional governments. E-mail: a.bonaccorsi@gmail.com

BORRELL-DAMIAN, Lidia. **Lidia** joined EUA in January 2006 and has been working within the areas of doctoral programmes and researchers' careers, university-industry collaborative research, knowledge transfer and also on the contribution of universities to regional innovation. Her main role is to help stimulate dialogue for action in order to develop a knowledge-based European society. She is also involved in developing international relations between Europe and Latin America in Higher Education and Research. Lidia Borrell Damian holds a Doctorate in Chemical Engineering from the University of Barcelona, where she was an assistant professor from 1990-1998. She was a post-doctoral researcher in the US (North Carolina State University) and Canada (The University of Western Ontario), from 1997 to 2000. Back in Barcelona, she worked for a private chemical company as deputy director for the R&D department. After two years of being immersed in the R&D industrial world, she moved to the public research management sector at the Pompeu Fabra University (Barcelona), where she was director of research services in 2003-2005, managing research projects, knowledge transfer activities and developing university research policy. E-mail: Lidia.Borrell-Damian@eua.be

EMPLIT, Philippe. **Philippe** is director of the Service OPERA-Photonics and Professor of Physics at the Université Libre de Bruxelles (ULB), in the Engineering, Science and Human Sciences faculties. He is the Director of all bachelor/master programmes of the Solvay Brussels School of Economics and Management (SBS-EM) and has acted during 2 years as the Vice-Chair of this business school. He presently is a member of the Task Force group of ULB, advising, on a scientific basis, the university authorities in their strategy versus rankings development. He also is a member of the Governing Board and of the Research Council of ULB, representing the academic staff. Philippe Emplit is a member of the Interuniversity Council CIUF of the Belgian Communauté française CFB, the organism responsible for collecting in a coherent way the statistics for all CFB universities. Besides his main research and teaching activity in the field of IT physics, experimental nonlinear guided wave photonics and ultrafast phenomena, he recently started, within its Business School activity, to study the methodology of assessment and ranking of universities and business schools, as well as some aspects of the so-called econophysics. E-mail: phemplit@ulb.ac.be

INZELT, Annamaria. **Dr Inzelt** is Founding Director of IKU Innovation Research Centre (1991). She is member of Economic Advisory Council of Budapest Business School, prime member of doctoral school at the University of Szeged and private professor at Budapest Corvinus University. She has been the first Hungarian representative in the OECD Working group of the National Experts of Science and Technology Indicators (NESTI) for 12 years. During her NESTI membership period she deeply involved in the revision of OECD Manuals and development of harmonised questionnaires. Beside that she chaired the revision of Hungarian STI statistics and introduction of Oslo Manual based innovation surveys. She is chairing an advisory group on modernising Hungarian STI indicator and analytical system. She was also involved in OECD collaboration with transition economies and with emerging, promising economies from different regions (ASEAN, African, and Latin-American countries). She is associate member of ERAWATCH network. On various national and international training seminars she is teaching on STI indicators, HRST mobility and university-industry collaboration. She is a scientific advisor to various EU FP projects. She has been participating and evaluating EU FP projects since 2004. She has been coordinating several EU sub-projects relating to new member states. Her main research interest includes the theoretical and practical issues of the innovation systems, the innovative capabilities and

performance of the different actors, business organisations and universities. She has well accumulated experiences in the field of science, technology and innovation measurement issues. She is studying the 'third mission' of universities. The challenges imposed by the internationalisation of R&D are studied by her from the point of view of host countries. She is interested in the subject of mobility and knowledge flows focusing on the higher educated people, the human resources of S&T and the impact of foreign direct investments on mobility. She is investigating the R&D and innovation policy in transition economies, and the changing rationales of policy making emphasizing the public-private interactions. In FP6 she was an active member of PRIME Network of Excellence. Beside several other activities she participated in PRIME Aquameth and Observatory of European Universities working groups and author and co-author of several papers and articles originated from these projects. E-mail: annamaria.inzelt@uni-corvinus.hu

MARKLUND, Goran. **Dr. Marklund** is Deputy Director General (Acting) at VINNOVA, which is the central Swedish Government Agency for funding mission-oriented R&D for innovation and sustainable growth. He is also Associate Professor in Economic History at Uppsala University with the focus on innovation and economic change. He has previously been Head of the Division for Strategy Development at VINNOVA, Science and Technology Attachée at the Swedish Embassy in Washington DC. He has also been guest researcher at the Center for International Technology Policy (CISTP) at George Washington University. Prof. Marklund often gives advice to the Government on research, innovation and growth policy issues. As a researcher, Prof. Marklund has primarily specialised in globalization, innovation and national competitiveness, with a particular focus on R&D, innovation and growth indicators. In this function he has closely followed OECD's and Eurostat's indicator work and often assisted at the meetings of OECD's group of national experts of science and technology indicators, NESTI. Prof. Marklund is currently chairman of the Advisory Board for R&D and Innovation Statistics at Statistics Sweden. E-mail: Goran.Marklund@VINNOVA.se

MOED, Henk. **Henk** is a senior staff member at the Centre for Science and Technology Studies (CWTS), in the Department (Faculty) of Social Sciences at Leiden University, as from 1986. He obtained a Ph.D. degree in Science Studies at the University of Leiden in 1989. He has been active in numerous research topics, including: the creation of bibliometric databases from raw data from Thomson Scientific's Web of Science and Elsevier's Scopus; analysis of inaccuracies in citation matching; assessment of the potentialities and pitfalls of journal impact factors; the development and application of science indicators for the measurement of research performance in the basic natural- and life sciences; the use of bibliometric indicators as a tool to assess peer review procedures; the development and application of performance indicators in social sciences and humanities; studies of the effects of 'Open Access' upon research impact and studies of patterns in 'usage' (downloading) behaviour of users of electronic scientific publication warehouses; studies of the effects of the use of bibliometric indicators upon scientific authors and journal publishers. He published over 50 research articles, and is editor of several journals in his field. He is a winner of the Derek de Solla Price Award in 1999. He published in 2005 a monograph, Citation Analysis in Research Evaluation (Springer, 346 pp.), which is one of the very few textbooks in the field. E-mail: moed@cwts.leidenuniv.nl

MUSTAJOKI, Arto. **Arto** is Professor of Russian Language at the University of Helsinki (UH). He has been involved in assessments and evaluations in different capacities. He chaired the Finnish Expert Group, which was responsible with two other universities for the Evaluation

of Uppsala University, the Evaluation Committee for Estonian Departments of Russian Philology and the International Evaluation Panel for the Faculty of Arts of Jyväskylä University. He has been a member of Steering Committees for several university level research exercises. He was Vice-Rector at the UH in 1992-1998 and chaired the Research Council for Culture and Society at the Academy of Finland 2001-2006. He belongs to the Steering Group of the European reference index for Humanities (ERIH). Arto Mustajoki took his master's degree in German philology and PhD in Russian language at the UH. His research focuses on contemporary Russian language, theory of functional grammar, cross-cultural interaction and modelling of (mis)communication. He has published six monographs and numerous articles on scientific topics and a book on research ethics and another book on university administration. E-mail: arto.mustajoki@helsinki.fi

NEWBY, Howard. **Sir Newby**, KB, CBE, BA, PhD, AcSS, is Vice-Chancellor of the University of Liverpool. Sir Howard was previously Vice-Chancellor of the University of the West of England (UWE) in Bristol having previously spent five years as the Chief Executive of the Higher Education Funding Council for England (HEFCE). He was Vice-Chancellor of the University of Southampton from 1994 to 2001 and was previously Chairman and Chief Executive of the Economic and Social Research Council (ESRC). From 1999 to 2001, Sir Howard was President of Universities UK, the UK body which represents the university sector. He was also President of the British Association for the Advancement of Science for 2001-2002. He was made a CBE in 1995 for services to social science and a knighthood in 2000 for services to higher education. Sir Howard was formerly a Professor of Sociology in both the UK and USA and has published a large number of books and articles on social change in rural England. E-mail: hnewby@liverpool.ac.uk

PEYRAUBE, Alain. **Alain** is currently Directeur de Recherche at the Centre National de la Recherche Scientifique (CNRS, Paris, France), Chair Professor of Chinese Linguistics at the Ecole des Hautes Etudes en Sciences Sociales (EHESS) and Scientific Director for the region "Ile-de-France" at the CNRS. He has been a corresponding member of the Academia Sinica of Taiwan since 2002, Adjunct Professor at the University of Science and Technology of Hong Kong since 2005, Honorary Professor at the University of Beijing since 2007, and member of the Academia Europaea (European Academy of Arts and Sciences) since 2006. He has been Founding Member of the Scientific Council of the European Research Council (ERC) (2006-2010), and member of the Standing Committee for the Humanities of the European Science Foundation (ESF) from 1998 to 2003, and again since 2007. He is Chair of the Steering Committee of the ERIH (European Reference Index for the Humanities), a project jointly sponsored by ESF and the European Commission ERA-NET programme "Humanities in the European Research Area" (HERA). He is member of the "Individuals, Societies, Cultures and Health" domain committee of COST (ESF and European Commission). His research interests are Chinese linguistics, typology of East-Asian languages, cognitive approaches to the diversity of languages, and the origin and evolution of languages. His latest research has been carried out within a broadly functional framework from a cross-linguistic perspective. He has published numerous articles and books. E-mail: Alain.Peyraube@ehess.fr

ROWLEY, Sue. **Professor Rowley** is Deputy Vice-Chancellor Research (DVCR) at the University of Technology, Sydney, Australia. Prior to her appointment in 2004, she was the Australian Research Council's Executive Director for Humanities and Creative Arts. In this role, she developed the ARC's inclusive approach to research in the creative arts, design and architecture. She was Chair of the Research Quality Framework (RQF) Creative Arts Panel in

its developmental phase. Sue's academic expertise is in contemporary Australian visual arts and design. She began her academic career at University of Wollongong in 1986 and was appointed UNSW Foundation Professor of Contemporary Australian Art and Head of the School of Art History and Theory in 1995. Complementing her written publications, she curated and managed national and international exhibitions of Australian craft and design, including multi-venue tours in the USA, Germany, Taiwan and Singapore. In the 1990s, she was President of Object, Australia's leading Centre for Craft and Design, and the Australian Council of University Art and Design Schools (ACUADS) and served on the Australia Council's National Infrastructure Committee. Sue's leadership roles build on 30 years' experience in community organisations (i.e. President of the Wollongong Women's Centre in the 1980s and Convener of the Richmond Education Centre, established by Labor's Schools Commission in the 1970s) and university bodies (i.e. elected staff representative on the Councils of Prahran Community College in the 1970s and University of Wollongong in the 1980s). Her current roles include membership of various boards (Capital Markets CRC, Australian Technology Park Innovations (ATPi), INTERSECT (NCRIS-funded NSW node for eResearch services and computing infrastructure), Sydney Institute of Marine Science (SIMS) and ARC Centre of Excellence in Ultrahigh Bandwidth Devices for Optical Systems (CUDOS). Sue is convener of the NSW D/PVCR Group and a member of the Executive of Universities Australia's D/PVCR Group. Email: Susan.Rowley@uts.edu.au

SALMI, Jamil. **Jamil**, a Moroccan education economist, is the World Bank's tertiary education coordinator. Mr. Salmi is the principal author of the Bank's new Tertiary Education Strategy entitled "Constructing Knowledge Societies: New Challenges for Tertiary Education". In the past sixteen years, Mr. Salmi has provided policy and technical advice on tertiary education reform to the governments of more than 40 countries in Europe, Asia, Africa, and South America. Mr. Salmi has also guided the strategic planning efforts of several public and private universities in China, Colombia, Kazakhstan, Kenya, Madagascar, Mexico and Peru. Mr. Salmi is a member of the Governing Board of the International Institute for Educational Planning, the International Rankings Expert Group, OECD's expert group on Assessing Higher Education Learning Outcomes, the International Advisory Network of the UK Leadership Foundation for Higher Education, and the Editorial Committee of OECD's Journal of Higher Education Management and Policy Mr. Salmi is a graduate of the French Grande Ecole ESSEC. He also holds a Master's degree in Public and International Affairs from the University of Pittsburgh (USA) and a Ph. D. in Development Studies from the University of Sussex (UK). He is the author of six books and numerous articles on higher education and development issues. His latest book, published in February 2009, addresses the *Challenge of Establishing World-Class Universities*. E-mail: Jsalmi@worldbank.org

SANCHEZ, Paloma. **Paloma** is a PhD in Economics and Professor of Applied Economics at the Autonomous University of Madrid (UAM). She is currently the Director of an Interuniversity Master and a PhD Programme on Economics and Management of Innovation and the Director of a recently created Chair UAM-ACCENTURE. She was the European co-ordinator of the MERITUM research project and the E*KNOW-NET thematic network, both on Intellectual capital issues, funded by the European Commission. She has participated in different evaluation and experts groups of the European Commission, the High Level Expert Group on "Reporting of Intellectual Capital for increasing investment in Research and Development" among them. She has advised the Spanish Ministry of Industry and the Spanish Ministries of Education and Science for 20 years and has represented the Spanish

Government at the Committee for Scientific and Technological Policy of the OECD for 15 years. She was the Chairperson of this Committee between 1990 and 1993. She is an Eisenhower Fellow since 1998. She has been asked to evaluate research activities of University teachers, first, as a member of the "National Committee to Assess Research Activities", in the Social Sciences and Humanities area, of the Spanish Ministry of Education (2003-2004), second, as a member of several evaluation committees appointed by different Spanish Universities. She has been Member of the Observatory of European Universities within the PRIME Network of Excellence. She is currently the director of a research project commissioned by the Madrid Regional Government to assess the results of the "Programas de Actividades de I+D entre Grupos de Investigadores de la Comunidad de Madrid. (Programme to encourage R&D activities among groups of researchers of the Madrid region). Her main research and publications are related to intellectual capital as a driver of innovation, economics of innovation, and innovation and technological policy. Over the last three years she has applied the intellectual capital framework to Universities and research institutions in order to help them to better measure, manage and report on their research activities. E-mail: mpaloma.sanchez@uam.es

8 Appendix II. Glossary

Term	Definition
Applied Research	Applied research is aimed at acquiring knowledge or understanding to meet a specific, recognized need. General examples of applied research would include using bacteria to inoculate plants against particular diseases, developing computer models of the atmosphere to improve weather forecasting, and examining the effects of the environment on human health or lifestyle.
Bibliometrics	A statistical or mathematical method for counting the number of academic publications, citations and authorship. It is frequently used to measure academic output. Data usually comes from an international database e.g. Thompson Reuters *Web of Science* or Elsevier-Scopus.
Blue-Sky Research	Often referred to as fundamental or basic research, blue-sky research is aimed at gaining more comprehensive knowledge or understanding of the subject under study, without specific applications in mind. Some general examples of basic research include research on the chemical properties of bacteria, analyses of the interaction of the means of temperature with the atmosphere, and investigation of neural pathways in the human brain.
Citation	A citation is a reference to a published or unpublished source which is formally counted as an indication of the scientific-scholarly impact of research.
Esteem	Recognition accorded to researchers by their peers for their achievements, leadership and contribution to their field(s) of research, expressed through, for example, the award of prizes and prestigious invitations such as international keynote addresses.
Expert Peer	Peer researcher with recognised expertise in the field of research under assessment.
Indicator	Indicators – also referred to as 'metrics' – measure the various components of research activity, including inputs, process, outputs, outcomes and impact and benefits.
Journal Impact Factor	A measure of the frequency with which the 'average article' in a journal has been cited in a particular year or period.
Knowledge Cluster	'Knowledge cluster' is the basic unit of assessment advocated by the AUBR Expert Group. A typical knowledge cluster comprises a group of researchers sharing a common field of investigation.
Knowledge Triangle	Refers to the interaction between education, research and innovation which are viewed as the key drivers of a knowledge-based society.

Multi- , Cross-, Trans and Inter-disciplinary Research	Research that employs the knowledge structures and characteristic behaviours of more than one discipline and may interrogate, critique and integrate specific disciplines and disciplinarity.
Practice-based Research	Practice-based Research aims to advance knowledge partly through, for example, the invention of ideas or prototypes, images, performances and artefacts, or clinical and class-room practice. It involves an integration of theory and practice to enhance understanding of the issues. It is commonly associated with medical, psychological or educational sciences, or the creative arts and media.
Productivity	The relationship between an output and the inputs of resources used to produce it; hence, a measure of the amount of output per unit of input.
Peer-reviewed Publication	A publication that has been submitted to referees for comment and recommendation in relation to its publication; normally refereeing involves peer-review. Also called a refereed publication.
Ranking	Rankings list items in a hierarchical order according to identified criteria. Rankings compare universities using weighted indicators which are aggregated, and then hierarchically ordered.
Rating	Ratings are a system of presenting the results of performance or quality assessment by grouping the results into different categories, e.g. A, B, C or *, **, ***. It is similar to banding but differs from ranking.
Scholarly Impact	Recognition of the originality and significance of the research by peers and its effect on the development of research in the same or related disciplines. This has traditionally been measured by citations.
Users and End-users	External non-academic groups which benefit from the uptake of research outcomes; an end-user may be a company, government, external agency, organisation or community.

9 Appendix III. Description of Indicators: Qualitative and Quantitative

9.1 RESEARCH PUBLICATIONS AND OUTPUTS

Description: Total Count of the Number of Publications and Other Outputs

Outputs are normally broken into the following categories:

- Books (authored research, critical editions, edited, major revisions)
- Book chapters
- Journal articles (refereed, scholarly journal, other contributions to scholarly reviewed journal, non-refereed articles, letters or notes)
- Reviews
- Conference publications (refereed, non-refereed, extracts of paper)
- Edited volumes of conference proceedings
- Audio-visual recordings
- Computer software, databases
- Technical drawings, designs or working models
- Design (major works in production or exhibition and/or award-winning)
- Patents or plant breeding rights
- Art (major works - exhibitions, performances, broadcasts, screened or published)
- Research reports - commissioned by government, industry or other
- Technical reports
- Legal cases
- Entries in a dictionary/encyclopaedia
- Maps
- Translations and editing of major works
- Case studies

Data collection is either undertaken by universities, or comes from an international database or increasingly from web-based technologies.

University-based data normally requires direct entry by researchers, often mediated through the Research Office. Data usually comes from an international database e.g. Thompson Reuters *Web of Science* or Elsevier-Scopus. There are also various citation indices, the most important of which are the *Science Citation Index Expanded*, the *Social Sciences Citation Index*, and the *Arts & Humanities Citation Index*. Increasingly web-based interfaces, such as Google Scholar, institutional repositories or other standardised web-based technologies, are used. In effect, researchers create and maintain online curricula vitae. In these cases, verification is undertaken by universities and auditing may also be used to exclude publications which do not meet criteria for inclusion.

Pro

Bibliometric data is collected on all research outputs in order to quantify the full extent of research activity. At national level, identifying categories for inclusion in data collection involves consultation with key discipline, research and university organisations and leaders, thereby increasing the legitimacy of the indicator. At this level, it is possible to narrow the

focus, and select only some aspects of research activity to be included. For example, international vs. national publications or visa-versa; research created for/directed towards end-users; research contributions to innovation and socio-economic benefit, e.g. research and technical reports, patents and plant breeder rights, computer software, designs and prototypes, and creative works.

It is possible to combine both ISI and Scopus products in research assessment. The Australia research quality assessment framework (ERA) will employ ISI products for Science disciplines and Scopus for its better coverage of the humanities and social sciences.

Con

The key impediment to bibliometric accounting is that different disciplines produce different types of research outputs which are not all easily recorded. Individual universities may be able to compile data on all categories of outputs, but this needs a high degree of compatibility for cross-institutional and cross-national comparability.

Traditional bibliometric practices systematically under-represents the research enterprise of the humanities, creative arts, architecture & design, and law. A further and increasingly significant consideration is that this indicator excludes research outputs that are created for, or directed towards, end-users.

International bibliometric databases focus primarily on journals that publish full text in English or at very least, their bibliographic information in English. They do not claim to have complete journal coverage, but rather to include the most important. ISI coverage tends to be excellent in physics, chemistry, molecular biology and biochemistry, biological sciences related to humans and clinical medicine; good, yet not excellent, in applied and engineering sciences, biological sciences related to animals and plants, geosciences, mathematics, economics, psychology and other social sciences related to medicine and health; moderate in other social sciences including sociology, political science, anthropology and educational sciences, and particularly deficient in most disciplines of the humanities including literature, arts, musicology, etc., and the creative arts and media.

A principal cause of non-excellent coverage is the importance of sources other than international journals, such as books and conference proceedings. In fields with a moderate ISI coverage, language or national barriers play a much greater role than they do in other domains of science and scholarship. In addition, research activities may be fragmented into distinct schools of thought, each with their own ‘paradigms’.

In short, in spite of its widespread and influential adoption, this indicator works well for some disciplines and poorly for others. It works well for disciplines, especially in the sciences, that are supported by coherent international research communities that have reached a reasonably broad consensus about the relative importance of the relevant journals. It works less well for newer and emerging disciplines, disciplines linked to professional practice, fragmented research communities, and research that are focussed on regional, national and local themes.

9.2 INDICATORS OF QUALITY AND SCHOLARLY IMPACT

9.2.1 CITATIONS

Description: Citations Data Derived from Standard Bibliometric Measures

Citation counts are manifestations of intellectual influence, as represented by the adage: 'standing on the shoulder of giants'. Citation impact is a quantitative concept that can be operationalised in elementary or more sophisticated ways, e.g. through crude citation counts or an advanced, normalised measure. Citation impact indicators may be considered 'objective' indicators in so far as they reflect properties of the cited documents themselves, they are replicable, and based on the practices and perceptions of large numbers of (citing) scientists rather than on those of a single individual scientist.

Data is purchased from commercial bibliometric providers, the most significant of which are Thomson Reuters and Scopus. Since 2004, Google Scholar has provided a freely-accessible Web search engine that indexes the full text of scholarly literature across an array of publishing formats and disciplines.

Some universities have established major research centres in bibliometrics, including the Centre for Science and Technology Studies (CWTS) at Leiden University and the Research Evaluation and Policy Project (REPP) at the Australian National University. These Centres use bibliometric data to undertake systematic evaluation and mapping of research at institutional, cross-institutional, national and international levels. For example, REPP regularly conducts bibliometric analyses of scientific publications produced under the patronage of publicly-funded bodies.

Pro

The prime factor in the widespread acceptance and use of this indicator is that it reflects good scholarly practice embedded in most disciplines, i.e. the attribution of intellectual advancements to those who made them.

Additionally the databases from which citations counts are assembled draw upon comprehensive publications sets (9,000 for Thomson Reuters; 18,000 for Scopus).

Con/Limitations

Citations are used as indicators of the extent of 'take-up' of research outcomes by scholarly communities but extraneous factors can also impact on the data, including:

- Publication language;
- Coherence of research communities;
- Citations practices of research communities;
- Extent to which citation practices coincide with accessible commercial bibliometric products;
- US-based and English-language journals, target readerships and content;
- Non-US local, regional or national content (histories, laws, policies, practices, etc).
- Coverage of a certain discipline in the data base used – citations might be found in publications not counted here.

There are three particular issues:

a) The journal impact factor

The impact factor of a journal in a particular year T is defined as the number of citations received in that year by all documents published in that journal in the two preceding years, divided by the number of citable documents published in the journal in the two preceding years.

The journal impact factor developed by Eugene Garfield and published by Thomson Reuters in the *Journal Citation Reports* (JCR) is probably the most widely dispersed bibliometric construct. It was used by Garfield to select the most important scientific journals for inclusion in the *Science Citation Index*. Their importance is assessed through a combination of an objective and truly unique internal monitor based on citation relationships among journals with assessments by experts from the various fields. One of the indicators applied in the internal monitor is nowadays known as the journal impact factor.

However, journal impact factors are inadequate measures of research performance because they are not always accurate; they can be manipulated, and are strongly affected by differences in citation practices among research fields. Thomson Reuters advises that:

> The impact factor should not be used without careful attention to the many phenomena that influence citation rates, as for example the average number of references cited in the average article. *The impact factor should not be used without informed peer review*.

b) Hirsch Index

The Hirsch index is based on the distribution of citations received by a given researcher's publications. A scientist has index h if h of [his/her] Np papers have at least h citations each, and the other (Np - h) papers have at most h citations each. In other words, a scholar with an index of h has published h papers, each of which has been cited by others at least h times. This appealing construct is highly biased towards 'older' researchers with long careers, and towards those active in fields with high citation frequencies, and provides an incomplete picture of a group's actual citation impact.

c) Normalised citation impact indicators

In an effort to overcome some of the problems identified above, a normalized or relative indicator of citation impact has been developed. This is the average citation rate of the papers published by a unit of assessment, divided by the world citation average in the scientific subfields in which the unit is active. This approach corrects for differences in citation practices among scientific subfields, for differences in the expected citation frequencies of the various types of papers (reviews tend to be cited more frequently than articles), and for differences in the 'age' of cited papers (older papers can be cited during a longer time period and therefore tend to have higher citation rates than more recently published ones).

9.3 PUBLICATIONS IN TOP-RANKED, HIGH-IMPACT JOURNALS

Description: The Number or Proportion of Journal Articles Published In Top-Ranked, High Impact Journals for the Fields of Research.

This indicator relies on the journal impact factor reports compiled by Thomson Reuters to identify top-ranked, high impact journals (see previous section).

However, noting the caveats aforementioned, universities, discipline organisations and government agencies may decide to develop their own rankings of journals – usually drawing on the journal impact factor reports. This exercise is often undertaken at the national level or by academies. For example, the Australian Research Council has undertaken a comprehensive ranking of journals which are then assigned to tiered bands, while the Norwegian Research Council is conducting a similar exercise based on the work carried out by NIFU STEP Institute. In this model, peers evaluate publication outlets and assign journals and book publishers to higher and lower categories. In a next step these categories are used to calculate indicators. ERIH (European Reference Index in the Humanities) is a project by the European Science Foundation to rank the top journals in 15 areas of the Humanities; it is available on the ESF website.

Pro

This indicator is a relatively recent innovation for systematic research assessment. It aims to counter the drive for high-volume publications counts with an incentive to publish in quality international journals.

The principal advantage is its wide acceptance by researchers, discipline and peak bodies, and universities of the ranking assigned to journals – that is, consensus that 'tier 1 journals' are indeed the most prestigious journals for each field.

Con/Limitations

In many fields, especially in social sciences and humanities, rankings of journals based on impact factors do not correspond very well to rankings made by experts on the basis of their perception on the status or quality of these journals. In those fields, one has to develop acceptable rankings, or at least rough categorizations, of journals and books using outcomes of surveys among peers. While feasible and most useful, this task involves a lot of additional effort.

9.4 OTHER INDICATORS OF PEER ESTEEM

Description: Peer Esteem Indicators Measure How Research and Researchers are recognised in the Wider Community.

Key indicators of peer esteem are:

- Presentation of keynote addresses at national and international conferences;
- Prestigious national and international awards and prizes;
- International visiting research appointments;

- Editorial and refereeing services for prestigious national and international journals/publishers;
- Election to learned societies;
- Curatorial/juried selection of work for exhibition, performance, recording, screening, etc.;
- Reviews of art works;
- Publicly commissioned artworks, compositions, film/theatrical scripts, etc.
- Appointments to advisory committees in national or international organizations;
- Appointments to research evaluation committees of individuals, groups, centre, etc.

This activity is normally recorded by direct entry by researchers.

Pro

These are widely used indicators of national and international leadership, influence, reputation and performance by relevant research communities and agencies. Normally, expert peer review or input is the basis of the award, election, invitation or appointment.

Con

These data are verifiable but there is no systematic verification protocol or technology available for verifying claims across diverse indicators – yet. Nor are there agreed equivalences and values that apply internationally and facilitate comparison across disciplines. A consensus is yet to be reached for the inclusion of prizes that cut across research and practice, such as the Nobel Prizes for Literature and Peace, and the Pritzker Prize in Architecture.

Reputation and leadership are not necessarily or entirely based on ground-breaking, paradigm-shifting, iconoclastic research. A researcher who has made a significant contribution to knowledge or challenged orthodoxy may find recognition and attribution a long time in coming. Research communities can also act as 'gatekeepers'.

Some research communities – astronomy for example – are coherent and international; others have not developed international organisations, awards and cultures. Prizes and other honours are a strong feature of science, technology and medicine research communities. They reinforce the influence of the key organisations that assess potential recipients and serve to promote research to governments and the broader community. Other research fields, notably in the Arts, Humanities and Social Sciences, are less well-organised and resourced, and less influential with government and industry, and have not developed awards and honours to this extent.

9.5 INDICATORS BASED ON HUMAN CAPITAL

Many of the indicators listed here can be interpreted as an indicator of performance but also as a precondition of performance.

9.5.1 DOCTORAL COMPLETIONS

Description: Number of PhD and Equivalent Research Doctorates and, as appropriate, Research Masters Degree Completions

Universities and, in some cases, government agencies, collect data for this indicator.

Pro

The widespread acceptance of the importance of research training is the key legitimating factor. PhD completions are regarded internationally as a common standard because the examination of PhD theses and submissions is undertaken by national and international expert peer researchers. Data is verifiable.

The use of this indicator promotes quality postgraduate supervision and programs. If resources are attached to performance against this indicator, universities, faculties and centres and institutes are likely to proactively promote quality control and institute programs and monitoring at the institutional level.

Con

Differences across disciplines as to the requirements for PhD research can affect the completion timetable.

This indicator is focused on quantity, e.g. the number of graduates. The quality of PhDs is probably not comparable across countries; it is not even comparable across universities within one country.

9.6 RESEARCH ACTIVE ACADEMICS

Description: Number or Equivalent Full-Time (FTE) of 'Research Active' Academics Employed by a University.

'Research active' is established by setting threshold levels of performance for a specific period (e.g. the previous year or three to six years) that normally would include:

- Number of publications and/or other research outputs;
- Number of research degrees supervised to successful completion;
- Research income;
- Number of funded research projects;
- Quantum of research income

For example, 'research active' might be defined as 1 publication and/or 1 PhD completion and/or 1 funded project and/or €50,000 income per assessment period. Some universities could require staff to achieve 2 or 3 out of these 4 indicators.

Universities collect data for this indicator.

Pro

This indicator is useful to drive improvement in research performance in universities and in internal units (faculties, schools, departments). Used as a proportion of total academic staff, this indicator can assist universities in building research capacity.

Con

The threshold level can be raised as stronger performance is achieved across the institution. This approach to indicators can also be drilled down differentially to faculties and other units within universities.

9.7 RESEARCH OUTPUT PER ACADEMIC STAFF

Description: Number of publications and other outputs per academic staff or Full Time Equivalent (FTE).

The total research output is divided by the respective number of staff. Who is counted depends on the type of output e.g. the number of PhD graduates should be related to the number of professors whereas the number of publications might be better related to the number of all research active staff.

Pro

This indicator provides a useful way to measure the contribution that individual active researchers make to the total university output. It can be adjusted to the size of a knowledge cluster or unit of assessment.

Con

Because universities are involved in a wide range of activities, not all academic staff may be research active. Therefore, it is may be difficult to determine the right denominator and to get reliable data on this.

9.8 NUMBER OF CO-PUBLICATIONS

Description: Number of Co-Publications within the Unit of Assessment or Collaborative Work with Researchers from other Universities/Organisations.

This indicator is important for identifying and supporting collaborative research work. It is based on a total count, best done by individuals and reported by the university.

Pro

Collaborative work with researchers in other universities, both nationally and internationally, shows the extent of a research engagement. The latter is an important indicator of internationalisation, which is itself a measure of peer esteem. Collaboration, measured by co-publications, within an institution may be an indicator for the integration of young academics in current research or of inter-disciplinarity. Collaboration with organisations outside the institution can be an important indicator of impact.

Con/Limitations

Co-authorship usually measures 'formal' collaboration, i.e. collaboration that leads to joint publications. However, collaborative research activity, especially with non-academic partners, is not easily reflected in the major international bibliometric data bases, especially work that may not be published as peer reviewed articles. When international indices are used for the analysis of co-authorship, the outcomes are reliable only if this index covers a field's publication output sufficiently well. In other words, it works better for the sciences than the arts, humanities or social sciences.

9.9 INDICATORS BASED ON INVESTMENT

9.9.1 EXTERNAL RESEARCH INCOME

Description: Level of funding attracted by researchers and universities from external sources.

This indicator measures competitive grants and research income from government, industry, business and community organisations. It is useful because it includes:

- Success in attracting grants in national and international competitive, and peer reviewed programmes;
- Overall level of financial support available to support research;
- Success in attracting funding and research contracts from end-user sources.

Universities usually collect data, although data may be provided by sponsoring organisations.

Pro

Research income is a useful indicator for measuring the scale of the research enterprise and its capacity to secure additional income through competitive grants and contact research, especially in science, technology and medicine. This indicator is comparable, and verifiable through audit, and can be useful for comparing research performance across the system and within universities. The willingness of industry to pay for research is a good lead indicator of value-for-money and usefulness, and its anticipated contribution to innovation and the economy. Research income can usefully be broken down by source of income, and level within the university, faculty and other units in order to make internal comparisons.

Competitive grants and funding are valuable indicators of past research performance and scholarly impact and can also be used to predict future performance. Funding from end-user sources (industry, the profession, governments and communities) is a good indicator of performance and contribution to innovation and/or social, economic, cultural, environmental benefit.

Con

Strictly speaking, research income is not an indicator of performance but rather is a precondition of performance, at least in science, technology and medicine fields.

Levels of external funding vary greatly across disciplines. Funding opportunities are less frequent and awards are significantly lower in the arts, humanities, and social sciences than in the sciences because the equipment and materials costs are often insignificant. For example, in countries where over half the total pool of funding is allocated to medical research, universities that do not have medical faculties will inevitably secure less funding.

It may be difficult to collect data from end users because this information may not be routinely collected by the Research Office. Again, there are differences between disciplines which need to be taken into account.

9.9.2 RESEARCH INCOME PER ACADEMIC STAFF

Description: Research Income Divided by the Number of Academic Staff or FTE.

Pro

This indicator enables assessment adjusted to the size of the knowledge cluster, and can be aggregated to institutional level. It supports cross-institutional comparisons adjusted for scale of institution.

Con

There can be difficulties determining comparability of the numerator, e.g. the number of academic staff or FTE because of the way different universities and countries consider this category. It may also be difficult to find comparable data on research income.

9.9.3 TOTAL R&D INVESTMENT

Description: Total Investment in University-Based Research from All Sources

This indicator includes all university allocations (e.g. investment) in research allocated from the government block or operating grants and externally-earned income, e.g. international student fees, training, commercialisation and philanthropic donations. Universities collect data for this indicator and self-nominate levels of funding for research from consolidated revenues.

Pro

Investment in research is arguably one of the strongest predictors of research performance at the level of the institution, region and country. This indicator links university R&D investment to Government and Business Investment in R&D (BERD) at regional and national level.

Con

It can be difficult to get valid, comparable institutional data, because a significant proportion of institutional investment is cross-institutional subsidisation.

9.9.4 RESEARCH INFRASTRUCTURE

Description: This indicator measures basic facilities that support research.

Infrastructure includes, inter alia, the library and digital access, certain laboratories and other research facilities – in other words, the environment the university provides for conducting research.

Pro

This indicator measures the research environment as a predicator of research capability and success. They can be a useful reference because it is difficult to characterise the strength of a research university without being able to gauge the quality of the infrastructure.

Con

National, historical, cultural and mission can make direct comparison difficult. It is often difficult to find quantifiable and comparable data or to express the existing facilities in terms of money. There are also differences between universities which are focused on the sciences rather than the arts and humanities. Furthermore, the move toward digital libraries could make comparison easier or more difficult.

9.10 INDICATORS OF ECONOMIC & SOCIAL BENEFITS

9.10.1 Commercialisation of Research-Generated Intellectual Property

Description: Typically, the indicator assembles data on:

•Invention Disclosures: the number of disclosures, indicated to the appropriate university office, e.g. Technology Transfer Office, of possible inventions to be considered for patenting;

•IP protected through patents (and plant breeder rights) – the number of patents applied for and/or granted

•IP licences, options and assignments (LOAs);

•Commercialisation revenues – income from licences, royalties etc.;

•Start-up companies;

•Value of equity in start-up companies;

•Number of jobs and average salary per job created by start-ups.

Universities usually regularly report on patents, LOAs, start-ups etc. Verification is possible (but complicated): income and equity data are included in audited financial reports and information on company ownership and current value is in the public domain.

Pro

This is an area of increasing significance to policy makers. The key legitimating factors are the link between IP commercialisation and economic benefit, the availability of data to support broad comparisons across national and international systems, and longitudinal analyses and the in-principle verifiability of the data.

Con

Patents are a very poor indicator. The number of patents filed differs across countries because of legislation or culture; for example, the European Patent Office has a centralised service for the 34 members of the European Patent Convention but there is no uniform system of enforcement. Each patent must be validated and enforced by each member state. There is also a difference in frequency; for example, up to 3 patents are often filed in Japan for 1 equivalent more encompassing patent in Europe; Technology Transfer Offices in the US are also more likely to file provisional patents as a matter of practice than their European counterparts. Patents only indicate that the patent agent was willing to file a patent believing that it was a unique idea – but this is not equivalent to it being a measure of commercial success.

The particular route to market implied by patents is relevant to some fields of science, technology and medicine but does not capture knowledge and technology transfer or translation achieved by other mechanisms. Moreover, it excludes the move towards open science or open source IT software or the embodied expertise of the social sciences.

9.10.2 EMPLOYABILITY OF PHD GRADUATES

Description: Industry Employment of PhD Graduates

Universities track the career destinations of their PhD graduates and alumni via post-graduation Career Destination Surveys and Alumni Databases.

Pro

This indicator attempts to measure the central role played by a highly trained and skilled workforce employed throughout society and the economy – the public and private sector –, and as leaders in innovation. The international consensus about the value, standard and relative consistency of PhDs makes this, potentially, a major indicator.

Con

Because the information is usually based on survey responses, the data can be unreliable. The information does not always specify the universities from which PhD holders graduated; hence their use for research assessment of universities is limited at this stage. Moreover, employability may be a factor external to the university, e.g. the state of economic development or point in the economic cycle. It is also difficult to align workforce information on PhD holders to the specific universities that trained them.

9.10.3 SCIENTIFIC PARTNERSHIPS AND COLLABORATIONS

Description: Number of Collaborative Projects with External Partners or Participation in Programmes Designed to Foster Collaboration.

This information is generally collated at the university level.

Pro

This is an important indicator for measuring scholarly involvement with other researchers and in turn the extent to which researchers are recognised by their peers, engagement with other organisations (e.g. public and private sector organisation, NGOs, and civil society), internationalisation engaged in by researchers, and the level of attractiveness by researchers in other universities and countries.

Con

The data is only in the early stages of being defined and collected. Many of the projects are conducted by researchers individually or privately, so the university's data may not complete. Results are often only internally published; the amount of money given is often not public.

9.11 END-USER ESTEEM

Description: Information Recognises Contributions to End-Users, the Economy and the Wider Society.

This indicator is the counterpart to OTHER INDICATORS OF PEER ESTEEM. They are similar to 8.4 above but seek to measure the contribution that research makes to society and the economy. The data can be collected quantitatively or qualitatively. The latter can be captured by involving key stakeholders and end-users directly in review panels or in written assessments. It includes:

- Appointments to relevant national or international organisations, committees and research councils;
- Policy recommendations;
- Requests for expert services;
- Invitations to sit on boards and/or management groups of commercial, government and/or not-for-profit organisations;
- Curatorial/juried selection of work for exhibition, performance, recording, screening, etc.;
- Critical review of art works; publicly commissioned artworks;
- Activities for public administration (committees, advisory roles);
- Standardisation bodies;
- Scientific communication (public understanding of science, conferences to the public, events, media coverage).

Normally this requires direct entry by researchers.

Pro

This indicator is used to measure leadership, influence, reputation and performance by relevant research communities and agencies. Normally end-user recognition and testimony is the basis of the award, invitation or appointment.

Con

Data is verifiable but there is no systematic verification protocol or technology available for verifying claims across diverse indicators – yet. In addition, there no agreed equivalences and values that apply internationally and facilitate comparison across disciplines.

10 Appendix IV. Case Studies of the Research Assessment Experience

10.1 AUSTRALIA

Executive Summary

Australia's Excellence in Research for Australia (ERA) will assess research quality within higher education institutions using a combination of indicators and expert review by committees comprising experienced, internationally-recognised experts. ERA outcomes will be reported by institution and by discipline, identifying areas that are internationally competitive and emerging areas for further investment. A trial of ERA in 2009 will evaluate two clusters, with the full ERA running in 2010.

Name/Title of Research Assessment Exercise:

Excellence in Research for Australia (ERA) Initiative

(including its predecessor, the aborted Research Quality Framework)

Policy Context, incl. circumstances under which the exercise came about:

The Minister for Innovation, Industry, Science and Research, the Hon. Senator Kim Carr, announced the ERA Initiative on 26 February 2008. ERA replaces the Research Quality Framework (RQF) developed by the previous Government over 2005-07 but not implemented prior to the Australian election in November 2007. The incoming Government aborted the controversial RQF on the basis that its design was cumbersome and lacked transparency, the costs of implementation (estimated to be $AUD100 million) would be too high, and it was not supported by influential groupings within the higher education sector.

Although it was abandoned, the ground-breaking RQF was on the threshold of implementation, with Panels appointed, Submission Specifications released in September 2007, and submission date set for April 2008. The exhaustive consultation, pilots, and technical and other development processes fed directly into the replacement ERA. Key elements of the Research Quality Framework were:

assessment of both the *quality* of research and its *impact* on the broader economy and society;

assessment based on evidence portfolios prepared by research groups and submitted by universities;

discipline-appropriate assessment by 13 discipline-grouping panels of expert peers, including research end-users and international researchers;

the quality of research to be demonstrated through 4 'best outputs' per researcher named in the research groups – with outputs to be stored in and accessed through digital repositories created by each university;

the impact on the broader economy and society of the Research Group's research to be demonstrated primarily through Case Studies;

the inclusion of 'Context Statements' describing the Research Group's:

history, strategic focus, research objectives, research income and main achievements;

involvement in collaborative research within the institution and the broader research environment (both within Australia and overseas);

support for early career researchers and postgraduate research students; and

evidence of scholarly esteem and reputation.

citations, number of publications in highly-ranked journals and other relevant indicators to be included in Evidence Portfolios to *inform* expert peer assessments;

reporting of separate ratings for quality and impact to be at the level of Research Group (not by individual researcher and not aggregated up to a unified university ranking);

anticipated use of RQF to determine allocation of significant funding for research through Government block grants programs that support research and research training in universities (the actual quantum and programs had not yet been announced).

ERA is intended to be differentiated from the RQF by its greater use of indicators and more 'streamlined' processes. Peer panels will not be required to read publications accessed via repositories but will rely instead on discipline-appropriate indicators. Eight ERA Panels will replace the 13 RQF Panels. Universities will report on research outputs by Field of Research codes.

ERA is currently under development by the Australian Research Council. The implementation will be rolled out over a period of 3-4 years commencing with Physics, Chemistry and Earth Sciences in 2008-9, followed by Humanities and Creative Arts (2009). No funding will be attached until a complete cycle has been trialled and broad support for the ERA Initiative has been achieved. The focus will be solely on assessing the quality of research – the attempt to incorporate an assessment of broader impact was dropped by the incoming Government.

The Australian Government intends that 'ERA will detail by institution and by discipline those areas that are internationally competitive, together with emerging areas where there are opportunities for development and further investment'.

Policy Objective(s):

ERA aims to:

identify excellence across the full spectrum of research activity;

compare Australia's university research effort against international benchmarks;

create incentives to improve the quality of research; and

identify emerging research areas and opportunities for further development.

The Government intends that ERA will command the confidence of the research community and, in particular, recognise on-going imperatives to:

promote collaboration between institutions and between university researchers and end users;

encourage scale and focus and thereby efficient use of research infrastructure and resources;

facilitate interdisciplinary research; and

minimise the burden on individual researchers, institutions and expert assessors.

Methodology, incl. time-frame, resources, costs, technologies:

ERA combines the use of research performance indicators with expert review by committees comprising experienced, internationally-recognised experts. Key elements of the ERA methodology are described below.

Discipline-specific Indicators are being, or will be, developed by Indicator Development Groups in the following categories:

Indicators of research activity and intensity

Indicators of research activity and intensity include research income, PhD completions and total number of research outputs. Submissions will include the number of publications over a 6-year reference period; the period for research income and PhD completions may be shorter.

Indicators of research quality

Indicators of research quality include analysis of publications and other research outputs using ranked outlets, citation analysis and percentile analysis where relevant. Currently, a comprehensive, 4-tier journal ranking index is being compiled with 17,000 journals have been ranked across 100 disciplines.

Indicators of applied research and translation of research outcomes

Indicators of applied research and translation of research outcomes will be determined at a discipline-specific level.

a Pilot ERA in late 2008, applying only to the Natural Sciences, to facilitate the consultation process and test the technology for online submission; no evaluation of research will be conducted.

The unit of evaluation is research disciplines within an institution, classified by the Australian and New Zealand Standard Research Classification (ANZSRC) Field of Research (FoR) codes at both two-digit (22 Divisions) and four-digit (157 Groups) level where relevant. ERA will assess *all* research produced by each discipline cluster within an institution during the reference period(s).

Cost estimates are not available currently.

Dissemination, incl. how much information is available regarding data and methods:

As with the previous RQF, ERA involves exhaustive consultation with researchers and the 39 universities in the Australian system. There is a strong requirement of procedural fairness and transparency and acceptance by key stakeholders.

The Government will report outcomes for each university by Fields of Research at the two-digit and four-digit levels, but not at the aggregated cluster level. It is not intending to generate an integrated ranking of Australian universities.

Intended and Unintended Consequences:

Intended consequences:

Following a complete cycle through the 8 panels over 3-4 years, the Government is likely to attach funding to outcomes. Some or all of the university block grants, currently based on performance-based indicators, for infrastructure, research training and research will be determined by ERA outcomes. Current indicators used are research

income, publications, postgraduate research degree completions and load.

Over time, increased confidence and investment in Australian research by State and Federal Governments and industry.

ERA is intended to drive Australian research to improve the quality and national/international reputation of research through:

Increased incentives for publication in influential, high-impact, international journals and other publication 'outlets'.

Greatly improved technical capacity for data collection across Australian universities.

Greater concentration of research funding in universities that are 'research intensive' (a category that tends to coincide with larger, older universities with strength in natural sciences and medicine).

Greater concentration of resources (funding, staffing, scholarships etc) within universities in areas of recognised research strength and strategic fit with university profiles.

Raised international standing for Australian research and increased international collaboration by researchers.

Unintended consequences:

Destabilisation and 'churn' in the system as the pressure to recruit talented staff results in rapid-paced mobility.

Increased differentiation amongst academic staff in terms of salaries and work-mix (teaching & research).

Increased emphasis on effective performance management of academic units (faculties etc), research groups and individual academic staff.

Greater cohesion and organisation of discipline groupings, especially in professional fields such as creative arts, architecture, law, business and commerce, but also in humanities, information & communication technology, and social sciences.

Potentially - reduced incentives (and disincentives) for collaborative research, especially cross-institutional research, and cross-disciplinary research.

Potentially - reduced incentives (and disincentives) for applied and collaborative research with industry and research focused on innovation and solving real world problems.

1. Observations /Additional Comments:

10.2 UNIVERSITÉ LIBRE DE BRUXELLES – BRUXELLES – BELGIUM

Executive Summary

In Belgium, assessment of university-based research has not yet been undertaken at national or regional / community level. However, in the context of internationalisation of higher education, the Université libre de Bruxelles (ULB) decided to initiate, in 2008-9, a process of systematic assessment of its research. The fields of research represented at ULB were divided by the university authorities into ten mutually exclusive disciplines, two of which are assessed per year. The research activities in a given discipline are assessed on the basis of a self-assessment report, prepared by all the research groups being evaluated, and an on-site visit of a panel of internationally recognised experts of the discipline in question.

Context:

"Evaluation de la recherche à l'Université libre de Bruxelles (ULB)"

(Université libre de Bruxelles Research Assessment Exercise – ULB RAE)

Policy Objective:

Goals of the exercise :

- to identify, and obtain a proper understanding of the University's research output and its evolution;
- to improve research performance and quality, and to achieve excellence;
- to develop a new internal managerial and governing tool for the university authorities and for the discipline-specific research teams.

Methodology, incl. time-frame, resources, costs, technologies:

Origin of the exercise : internal

Type of assessment : peer review of discipline-specific research teams (not individuals)

Time frame: the assessment is carried out every 5 years (for each discipline)

Timetable of the process for each discipline :

•research teams active in the discipline in question are identified, and each team identified nominates a team leader, who will be in charge of all contacts (with the university and the peer-review panel) during the assessment process;

•an academic coordinator is nominated by the university authorities (usually from outside the country);

•a panel of 10 experts is convened by the academic coordinator;

•each team has to prepare a self-evaluation document, based on a common template :

1. Presentation of the team

1.1. Introduction to the general research topics (2 pages max.)

1.2. Presentation of the most important research results (2 pages max)

1.3. Outlook on future research activities envisaged (1 page max.)

1.4. SWOT analysis (2 pages max.)

1.5. Scientific and/or societal impact of the research activity / activities undertaken (2 pages max.)

1.6. Short CVs (1 page max.) of senior scientists (permanent academic staff), based on a common format;

1.7. Representative publications authored by members of the team (and justification of the choice made)

2. Scientific activities of the team

2.1. Publication strategy (1 page max.)

2.2. List of publications

2.3. List and synthetic description of research projects

2.4. List of other research outputs

3. Resources of the team

3.1. Staff

3.2. Teaching activities (incl. size of the classes)

3.3. Financial data

3.4. Third-mission activities

- the documents prepared by the disciplinary teams are compiled by the Research Department and sent to the expert panel;
- based on the analysis of this compilation, each expert assesses each team on seven indicators, giving a grade for each indicator (between 1 and 10); the seven indicators are: global evaluation, expertise of the reviewer in the field of the team, novelty of the research, quality of the research outputs, quality of the team's research strategy, scientific-scholarly impact of the research activity, societal impact of the research activity;
- all expert evaluations are transmitted to the team leaders;
- a one-day on-site panel meeting is organized, attended by the 10 members of the panel, the team leaders, and several members of the Research Department;
- two assessment reports are written by the coordinator and the members of the panel :

1. General report on the discipline ("public")

1.1. Context

1.2. The discipline and its teams

1.3. The assessment methodology

1.4. Conclusions of the members of the panel

1.5 Comments on the process of assessment

1.6. Observations of the academic coordinator

Annexes : data collection on the teams, evaluation files, CVs of the members of the panel

2. Report on the teams ("private")

2.1. The team

2.2. The team's research activities, including an assessment of the degree of originality of the activities undertaken

2.3. The research outputs and their quality

2.4. The management and the research strategy of the team

2.5. Conclusions and recommendations

2.6. International benchmarking of the team within its discipline

Dissemination, incl. how much information is available regarding data and methods

All documents concerning the ULB Research Assessment Exercise (goals, guidelines, etc.) are publicly available.

The general reports on disciplines are "semi-public", i.e. internally available for ULB members of staff.

The reports on discipline-specific teams are strictly "confidential", i.e. only available to ULB authorities and the team leaders concerned.

Intended and Unintended Consequences:

No consequence at the moment ; the first cycle of the exercise has not yet been completed.

Observations:

The first cycle of the ULB Research Assessment Exercise is about to end for the first two disciplines evaluated (physics, economics & management) (September 2009).

Additional Comments:

The ULB Research Assessment Exercise is similar to the system in operation since 1996 at ULB's sister university, the Vrije Universiteit Brussel (VUB), and described by N. Rons et al. in *Research Evaluation*, Volume 17, Number 1, March 2008 , pp. 45-57.

10.3 FINLAND (AALTO UNIVERSITY)

Executive Summary

The research evaluation included some innovative ways of using the peer-review method. The panels were asked to give their written statements and quantitative grades not only on research quality, but also on scientific-scholarly impact, societal impact, the research environment and the future potential of the unit in question.

Name/Title of Research Assessment Exercise:

Research Evaluation of Aalto University 2009 (AALTO Evaluation)

Policy Context, incl. circumstances under which the exercise came about:

The Helsinki University of Technology (TKK), the Helsinki School of Economics (HSE), and the University of Art and Design Helsinki (TaiK) are in the process of merging. The new university will be called Aalto University. The idea underlying the merger is to create a world-class research university. Aalto University is to become operational in August 2009.

Policy Objective(s):

According to the Charter of Foundation, Aalto University's activities are based on top-level research. Thus, conducting a research evaluation even before the new university started its work was a logical decision. As a result of the evaluation, senior management will know in which research areas Aalto University achieves the best results or has the potential for reaching the highest international level, and in which areas additional support is needed. At the same time, the evaluation will provide a benchmark for further development of research..

Methodology, incl. time-frame, resources, costs, technologies:

The overall procedure of the evaluation was rather traditional:

Thorough planning of the evaluation process by the Steering Committee.

Data collection and self-evaluations of departments according to strict rules laid down by the Steering Committee.

Evaluation, including one-week site visits, undertaken by nine international panels (all panels visiting the University at the same time).

The panels were asked to prepare written assessments and to give grades (from 1 to 5) on the basis on the material submitted by the departments and of interviews conducted during their site visits. The following evaluation criteria were used:

Scientific Quality. The evaluation covered publications and other research outputs including artistic or design productions. The peers were asked go give equal weight to all research, whether basic or applied. Additionally, they were informed that "in some fields it may be appropriate to publish rather extensively in Finnish and/or Swedish, which are the official languages of Finland. The language of publications should not have a negative effect on the rating as such: also publications in Finnish or Swedish may be substantially at the Outstanding International Level."

Scientific Impact. Such questions as: Are research outputs of the Unit published in the best forums in the field and do they have a notable impact on the development of the field? Do the members of the Unit occupy important positions in influential academic and professional associations in the field? Are they sought-after experts in tenure committees, chair appointments, research assessments and are they regularly invited to speak at the most important conferences in the field? Are the doctoral graduates of the Unit hired by the leading universities across the globe? To which extent does the Unit participate in international research networks and projects? Amount of external (basic) research funding e.g. from the Academy of Finland.

Societal Impact. The relevant indicators of societal impact include expert tasks, popularised works, media visibility, external funding relating to research cooperation with non-academic institutions (especially TEKES and EU funding), cooperation with the public and private sector outside academia, patents, start-up companies etc. Due to the scope of the evaluation the societal impact of teaching as well as the societal impact of the production of Bachelor- and Master-level graduates fell outside the scope of this assessment.

Research environment. The panels were asked to assess issues such as research leadership, research strategy, including human resources strategy and the focus of research, as well as issues such as the availability and quality of support services, research infrastructure, databanks, the teaching load of research-active staff, and the staff-student ratio.

Future Potential. While it is impossible to assess the future, the panels were asked to focus on such indicators of future research potential as the Unit's vision and plans for the future, how realistically the Unit saw its strengths and weaknesses, opportunities and threats, and whether the Unit had a carefully contemplated plan for managing such factors. Also, issues such as the age and career profile of the research-active staff, the size of the unit, and the ability to attract high-quality international doctoral students and researchers played a role here. Other important indicators of future research potential included the ability to secure competitive funding, the capacity to focus the Unit's research on topical issues, and the departments involvement in promising international collaboration networks.

Dissemination, incl. how much information is available regarding data and methods:

The Evaluation has its own web site with all the necessary information in English

http://www.aaltoyliopisto.info/en/view/innovaatioyliopisto-info/research-evaluation

Intended and Unintended Consequences:

It is too early to think about the consequences.

Observations /Additional Comments:

The Units of Evaluation (departments) were rather had some difficulty in doing their self-evaluations, especially in answering questions regarding issues which are not usually raised in research assessments (societal impact, research environment, future potential). One may, however, hope that compiling the documents urged them to think about these important things.

10.4 FINLAND (HELSINKI UNIVERSITY)

Executive Summary

The case study describes the research assessment exercise carried out by Helsinki University for its own purposes. The exercise has the following characteristic features: 1) departments as assessment units; 2) strict guidelines for collecting background data; 3) expert panels comprised of eminent foreign scholars / scientists; 4) site-visits of one week duration; 5) clear preference for quality over quantity of publications.

Name/Title of Research Assessment Exercise:

Research Assessment Exercise of the University of Helsinki 2005 (UH RAE)

Policy Context, incl. circumstances under which the exercise came about:

In Finland several types of university assessments are in place:

The Academy of Finland carries out international assessments of research fields at irregular intervals. All of them are published on the website of the Academy of Finland. The evaluation report for biotechnology can serve as an example: http://www.aka.fi/Tiedostot/Tiedostot/Julkaisut/Biotechnology%20in%20Finland.pdf

The Finnish Higher Education Evaluation Council (FINHEEC, an independent expert body nominated by the Ministry of Education) provides evaluations of education and teaching and in this context identifies, at three-year intervals, "high quality educational university units", which are awarded substantial additional resources. FINHEEC is also responsible for auditing procedures for universities and polytechnics.

A number of individual universities carry out research assessments, and evaluations of teaching and education on their own initiative. The first research assessment of this kind was carried out by the University of Helsinki in 1999. Several other Finnish universities have since organized similar exercises. Their purpose varies: to enhance research within the university in question; to gain visibility in the media and the opportunity to draw attention to the high level of research; to acquaint foreign researchers (experts serving on evaluation panels) with the research carried out and the researchers working at the university in question.

Policy Objective(s):

Helsinki University is one the founders of the League of European Research Universities. As part of its strategy, it regards research and the training of researchers as its main objectives. Regular research assessments are part and parcel of this strategy. The rather high expenditure on the RAE and substantial financial incentives provided on the basis of RAE results reflect the high priority awarded to research at HU.

Methodology, incl. time-frame, resources, costs, technologies:

The main target of assessment

The quality of the research carried out by the various departments and by a number of independent research institutes (individual researchers or research groups are not

evaluated).

An additional target of assessment: comments (without grades) were also requested on the departments' activities in interaction with society.

Methodology:

In the latest RAE, each department provided a large self-evaluation report following strict guidelines: list of research-active staff, description of the department's research profile (2 pages max.), complete list of publications, list of selected publications (2 per research-active staff) plus two copies of the publications selected, a short list of the best publications (three per permanent professorship in the subject area in question), list of doctoral theses, implementation of the ´99 recommendations (2 pages max.), SWOT analysis on research activities (2 pages max.), description of interaction between the unit and society (2 pages max.), other academic activities (visits abroad, list of academics visiting the unit, number of invited guest speakers, and editorial assignments), budget and external funding

The assessment was conducted by 21 international peer review panels (altogether 148 experts: 83% from Europe; 9% from Finland: 12% took part in the previous RAE: 30% women).

Panel members received all relevant materials ahead of the site visit, including information on the Finnish university system and research policy. They then spent one working week in Helsinki, writing the Evaluation Report before leaving; the Panel Chair finalized it afterwards.

For quantitative evaluation a scale of 7 to 1 was used.

Schedule:

Deadline for submitting of evaluation materials from departments: 28 February 2005.

Site visits of the panels: May-November 2005.

Publication of results on the university web site: 1 March 2006 (Summary Report, Individual Evaluation Reports).

Organization:

Steering Committee (four persons) chaired by the vice-rector.

For practical matters, an Evaluation Office was established: a Coordinator (PhD) from April 2004 to March 2006, two Planning Officers (altogether 20 months), and two Project Secretaries (altogether 17 months). All the staff of the Evaluation Office were recruited through an open application procedure.

Costs:

Salary costs of the Evaluation Office: 320 000 euro; peer review costs: fees 256 000, travel and accommodation 190 000 euro; other costs 130 000 euro.

Dissemination, incl. how much information is available regarding data and methods:

All materials of UH RAE (Terms of reference, Guidelines for the departments, evaluation report, etc.) are available in English on the following web site:

http://www.helsinki.fi/research2005/english/index.htm

Intended and Unintended Consequences:

There is no denying that there may be a bias problem in an assessment in which peers evaluate and grade colleagues from their own field. Panellists are aware that their judgment will have a negative or positive impact on the allocation of resources to their own fields at the university concerned. This may lead to an inflation of top grades. Such consequences can be avoided by providing external experts with precise descriptions of the different grades; this method was in fact used. Nevertheless, the problem exists. Perhaps the only way to obtain more reliable quantitative assessment results is to compare the outputs of departments with those of similar units at other universities.

10.5 FRANCE

Executive Summary

The case study describes the research assessment exercises undertaken by the *Agence d'evaluation de la recherche et de l'enseignement superieur* (AERES). AERES has three sections: one section is concered with the evaluation of higher education (including universities) and research institutions as a whole, the second one with the research undertaken in these institutions, and the third one with education and training. The evaluation of research is based on peer-review methodology.

Name/Title of Research Assessment Exercise:

AERES (Agence d'evaluation de la recherche et de l'enseignement superieur)

[National Agency for Evaluation of Research and Higher Education]

Policy Context, incl. circumstances under which the exercise came about:

The French Ministry of Higher Education and Research established the AERES National Agency two years ago to undertake, at national level, the evaluation of research in both research organisaitons and universities. The organisations (National Committee of the CNRS and the National Council of Universities for Higher Education Institutions) which used to be responsible for the evaluation have not been disbanded. However, they are not now allowed to evaluate research any more, only retaining their responsibility for hiring new faculty (called 'enseignants-chercheurs') and full research fellows ('chercheurs'), and deciding upon applications for promotion.

Policy Objective(s):

The main objective of AERES is the evaluation of all French research teams and research centres, i.e. units supported by CNRS or other research organisations (INSERM, CEA, IRD, INRA, etc.) and by one of France's 85 universities, as well as the units only supported by the universities, and not linked to the CNRS etc.

The evaluation committees convened by AERES are asked to identify excellence among the teams and to compare teams active in the same field of research by ranking them. Each team is usually given one of the following grades: A+, A, B and C.

Methodology, incl. time-frame, resources, costs, technologies:

Peer-review methodology.

The evaluation committees (or panels) are comprised of seven to ten members each; each committee has a chair. Committee members are appointed by the coordinators of the different scientific fields at AERES. Non-French members may also act as chairs. The committees visit the research units to be evaluated for a full day (or two days in the case of big units).

The main criteria used in the evaluation are as follows (from more important to less important):

Quality of research (mainly based on the scientific production of the lab);

International recognition of the lab (by using some quantitative indicators, such as the number of citations etc.);

Participation in national and/or international research networks or programmes; research grants obtained from ANR (Agence Nationale de la Recherche) at a national level, from the European Commission or from other European or international funding organisations;

Risk-taking in research, and openness to interdisciplinary interfaces (frontier research);

Openness towards societal challenges;

1. Investment in the dissemination of scientific culture; valorization (patents);

2. Number of full research fellows (chercheurs) in the lab, compared to other members (university faculty) who teach and are active in research for 50% of their time (enseignants-chercheurs);

3. In some domains (mathematics, for example), the supervision of PhDs is also taken into consideration in evaluating research.

As regards the evaluation of the scientific-scholarly outputs (the main criterion), there are different requirements in different disciplines: for every four-year period, in Mathematics, a *chercheur* or an *enseignant-chercheur* needs to have published 2 papers in an international scientific outlet ranked A; in Physics, Chemistry or Sciences of the Universe, 4 papers for a *chercheur* and 2 papers for an *enseignant-chercheur*; in Engineering sciences, and in Sciences and Information and Communication´Technologies, 3 papers for a *chercheur and* 2 for an *enseignant-chercheur*; in Life Sciences and Medicine, 4 papers for a *chercheur* and 2 for an *enseignant-chercheur*; in Humanities and Social Sciences, 4 papers for a *chercheur* and 2 papers for an *enseignant-chercheur*.

Dissemination, incl. how much information is available regarding data and methods:

Which outlets are considered ranked international scientific outlets is explained in detail on the AERES website. Apart from the best journals in each discipline, monographs and chapters in books are taken into consideration in the Humanities and Social Sciences, and papers published in proceedings of international conferences for the Engineering sciences.

The evaluation reports are sent to the research teams and to the presidents of the universities, who have the opportunity to comment on the evaluation reports and discuss them with those in charge at AERES before the reports are finalised.

Intended and Unintended Consequences:

The main consequence of this evaluation system is the allocation of funding to the research teams by the Ministry of Higher Education and Research and by the research organisations.

More resources are given to teams rated A than to teams rated B or C.

Another obvious consequence (intended or perhaps unintended) is that some institutions use the exercise as a benchmarking exercise in an attempt to improve their rating by closing down sectors which show a poor performance.

Observations /Additional Comments:

10.6 GERMANY - FORSCHUNGSRATING (CONDUCTED BY WISSENSCHAFTSRAT

Executive Summary

The German Science Council Rating carried out a pilot study based on peer review, information from departments, metrics and a reviewers' panel. The pilot did not include on-site visits. In a multi-step assessment process, the institutions were first evaluated by at least two experts independently before each rating was discussed in plenary sessions. In the final reports, peer esteems for five dimensions are published.

Name/Title of Research Assessment Exercise:

Forschungsrating (conducted by Wissenschaftsrat, pilot study)

Policy Context, incl. circumstances under which the exercise came about:

At the request of the German federal government and the state (Länder) governments, the Wissenschaftsrat (German Science and Humanities Council) has analyzed the methods and use of rankings in the higher education and research system. In its Recommendations for rankings in the system of higher education and research, the Council has established standards of good practice for rankings. In view of the importance of research activities for the success and international renown of scientific institutions, and considering the unsatisfactory state of comparative assessment of research performance in Germany, the Council has in addition developed a concept for a new research rating process.

Policy Objective(s):

In summer 2005, the Council decided to conduct a pilot study. In order to test the feasibility and usefulness of the research rating process across a range of methodically diverse subjects, chemistry and sociology were chosen for the pilot study. The Council secured the support of the relevant learned societies, i.e. the German Chemical Society and the German Sociological Association.

Methodology, incl. time-frame, resources, costs, technologies:

The research rating is distinguished from popular media rankings and league tables by a number of unique characteristics:

- Research quality is assessed by informed peer review on the basis of an extensive analysis of quantitative and qualitative data.
- Criteria and data are defined in a discipline-specific manner by experts from the individual fields of research.
- Research quality is assessed at the level of research units, making it possible to identify differences in research quality within individual institutions.
- The institutions are assessed by six different criteria which are not aggregated to an overall result. Thereby, the assessments reflect the institutions' different profiles and missions.

- For the first time, universities and non-university research institutions in Germany are evaluated in a single, comprehensive exercise.

In a multi-step assessment process, the institutions were first evaluated by at least two experts independently before each rating was discussed in plenary sessions. For the rating a five point scale ranging from "excellent" to "not satisfactory" was used.

- Dimensions and criteria of the assessment:

- Research (Quality of research, Impact/effectiveness, efficiency)
- Promotion of young researchers
- Knowledge transfer (Transfer to society, knowledge dispersion)

Indicators:

- Quality of research: Citations per paper, raw and standardised; number of publications; list of publications; other research products, e.g. databases, software; third-stream projects; prizes (chemistry: work-done-at principle; sociology: current-potential principle; different data bases used for the two disciplines)
- Impact/effectiveness: Number of publications; number of patents; third-stream funding in total; percentage of externally funded personnel; number of citations; guest researchers; self-report on interdisciplinarity; citations in other disciplines.
- Efficiency: "Impact" relative to human resources; publications/citations/third-stream funding/patents per full-time equivalent researcher.
- Promotion of young researchers: number of people working for their doctorate; number of (female) doctorates; PhD programmes; PhD process; number of postdoctoral grants; prizes to junior researchers.
- Transfer to society: number of (licensed) patents; money gained by licenses; funding by industry/business; spin-offs; consultancy work.
- Knowledge dispersion: Number of vocational trainings finished; lifelong learning; examples of spreading knowledge outside the scientific community.

Time-frame: November 2005 – December 2007

Costs: 1.1 million euro in total. 15 to 16 peers per discipline, estimated work load about 4 -5 (chemistry) and 8 - 10 weeks (sociology) respectively.

Dissemination, incl. how much information is available regarding data and methods:

The final results from the pilot study were published in December 2007 (chemistry) and April 2008 (sociology). They are summarized in the following documents:

The full results are only available in German.

Both expert groups and the steering committee have submitted reports on the pilot study to the Science and Humanities Council. The Council discussed the results of the pilot study during its 2008 May session, recommending that the research rating be continuted with two more subjects in order to further improve its methodology. Details were to be announced in fall 2008.

Intended and Unintended Consequences:

Observations /Additional Comments:

Only the peer esteems for the five dimensions were published, the indicators were not. Furthermore, different indicators were used for one and the same dimension; no details on these were provided. The relation between performance judged on the basis of indicators and the marks given by the peers is, therefore, not transparent. It is difficult if not impossible to compare the outcomes of the *Forschungsrating* with the outcomes of assessments based on indicators.

10.7 GERMANY - CHE UNIVERSITYRANKING, CHE RESEARCHRANKING

Executive Summary

CHE Rankings is a discipline-specific, multidimensional system aimed at providing information for students. Its main emphasis is on teaching, where national comparisons are made. Most prospective students use it as one source of information when deciding on where to study. CHE has another, research-focused system, which is not comprehensive, but universities do use it to benchmark performance. Both rankings have helped to make it clear that there are differences between German universities.

Name/Title of Research Assessment Exercise:

CHE UniversityRanking, CHE ResearchRanking

Policy Context, incl. circumstances under which the exercise came about:

In the late 1980s, the first university rankings were set up by newspapers.

Their methods were not at all adequate; one of the founding tasks of the CHE was to construct a ranking based on a sound methodology and involving the universities in the development process. The CHE is an independent non-profit organisation. It was founded in 1994 by the Bertelsmann Foundation and the German Rectors' Conference as a non-profit limited company. It decides independently on the choice of its topics and projects as well as on their realisation.

Policy Objective(s):

The CHE University Ranking is designed to help prospective students make an informed choice of study program and university, while the CHE Research Ranking of German Universities addresses scientific communities or universities The latter is designed to identify research-active departments at German universities in specific disciplines. It may also be used as a benchmarking instrument for universities. Both rankings are meant to develop transparency of university performance in different fields and to stimulate competition between universities.

Methodology, incl. time-frame, resources, costs, technologies:

The CHE Rankings provide an indicator-based, multidimensional system. Both rankings are discipline specific, covering about 25 disciplines (CHE Research Ranking: 15 disciplines), each of which are revisited every three years. The indicators are based on different data sources: data collected directly at the universities, publication databases (Web of Science and national databases of scientific-scholarly publications) and a survey conducted among professors. The main characteristics of the CHE Rankings are the following:

• no aggregation of indicators across the whole of a university, but subject-specific data analysis and presentation of results.

• no weighted or non-weighted total value for the research performance of a given department, but examination of different indicators in a multidimensional ranking

• no individual rank positions, but profiles of universities

Indicators used for research assessment:

- Amount of third stream funding (per researcher): Different sources of third stream funding may be reported by the universities; no distinction is made between different sources. Categories used include the German Research Council, EU projects, money from state governments or the federal government, and from industry or foundations.
- Number of publications (per potential author): For the different disciplines different approaches are required, depending on the publication habits in the various subject areas. Not all publications of each department or institute are taken into consideration, but only a certain sample which is selected on the basis of the data base(s) used, the authors' names and the time window used. If very heterogeneous data bases are used, a weighting scheme is applied taking into account the number of pages and of authors. In very few cases a set of core journals was established to identify important publications, which were then given more weight. The CHE uses the "current potential" principle, which means that publications of all authors working in a certain department at a certain date are taken into consideration, and not all publications produced in the department over the past few years.
- Citations per paper: If the Web of Science is used, a citation analysis is carried out as well; the indicator is not weighted differently for different fields.
- Number of patents (per researcher): Since 2006, in Germany, all inventions made by researchers at universities are in the first instance owned by the universities, and not by the inventors themselves. Every researcher who wants to have an invention patented first has to inform his/her university, and only if the university refuses to have the invention patented, can the researcher do so on his/her own. This regulation makes it possible to ask for the number of inventions registered with the universities offices dealing with the transfer of knowledge and technologies. In the given case, CHE requested information about all the inventions reported to the universities in the previous three years by researchers from the fields included in the Ranking.
- Number of PhDs: In the institutional survey, universities were asked for the number of doctorates completed in their departments over the previous three years.

Time frame: About one year for one cycle; in each cycle, data are collected from the previous three years.

The ranking team is comprised of six CHE members of staff; additional human resources are provided by *Die Zeit* for the publication and programming of the online-version. There are additional costs for the poll conducted among students and for the analyses of publications. The universities contribute to the exercise by completing the institutional questionnaires and helping with the logistics of the student survey.

Dissemination, incl. how much information is available regarding data and methods:

The results of the CHE University Ranking are published in "Der Studienführer" once a year; the indicators are freely available online and may be used interactively. The CHE Research Ranking is published as a pdf download, containing all indicators and additional figures. A comprehensive paper on the methods used in the two rankings is available on the CHE

website, as are the questionnaires used.

Intended and Unintended Consequences:

Intended consequences

The ranking has had some influence on the choice of university made by specific groups of students; most prospective students use it as one of a number of sources of information when deciding where to study. The ranking helped to make it clear that there are differences between universities.

Unintended consequences

The quality of the data is improving because of public pressure Since the publication of the first CHE Research Ranking, self-awareness and self-assessment within disciplines / departments and institution have considerably improved. At the same time, universities have begun to provide data in an "intelligent" way in order improve the position of their institutions in the Rankings.

Initially, the assessment of research output was distorted by the fact that only publications authored by professors were selected for analysis. This has since been changed; CHE now takes all publications into consideration that were produced by members of staff holding a doctorate.

Observations /Additional Comments:

10.8 GERMANY - INITIATIVE FOR EXCELLENCE

Executive Summary

The Initiative for Excellence has aroused some controversy in Germany because it is placed somewhere in the middle between evaluation and application for funding, and because there is no information available about the indicators used. The results of the Initiative were based on the assessment universities' plans for the future undertaken by an international panel. The Initiative was designed to identify excellence and allocate funding.

Name/Title of Research Assessment Exercise:

Initiative for Excellence of the German Federal Government and the state governments

Policy Context, incl. circumstances under which the exercise came about:

The Initiative for Excellence of the German Federal Government and the state governments aims at promoting excellent research at German universities.

Policy Objective(s):

The Initiative for Excellence aims at

- supporting top-level university-based research and improving its international visibility
- creating outstanding conditions for young scientists at universities
- deepening cooperation between disciplines and institutions
- strengthening international cooperation in research
- promoting equal opportunities for men and women
- intensifying scientific and academic competition and generally improving scientific and academic standards in Germany.

Methodology, incl. time-frame, resources, costs, technologies:

The Initiative for Excellence is administered by the German Research Foundation (DFG) together with the German Council of Science and Humanities (WR). It includes three lines of funding:

1. Graduate Schools for the promotion of young researchers
2. Clusters of Excellence for the promotion of top-level research
3. Institutional Strategies for advancing top-level university research.

To date, the Initiative for Excellence has been implemented in two rounds of funding: in 2005/2006 and in 2006/2007. Each round of funding had a pre-selection and a final selection round. In the pre-selection round, universities submitted Draft Proposals. These were reviewed by internationally appointed panels of experts. The reviews of the Graduate Schools and the Clusters were discussed in the 'Expert Commission', appointed by DFG; those for the Institutional Strategies in the 'Strategic Commission', appointed by the German Council of Science and Humanities. Both formed a 'Joint Commission' for the pre-

selection round. The universities chosen from this stage subsequently presented their full proposals. These were assessed in an identical procedure. The 'Grants Committee' made up of the Joint Commission and the federal and state ministers of science and research then agreed on the proposals to be funded.

In the first round of funding, 319 Draft Proposals were submitted by 74 universities. Of these, 90 propsals (39 Graduate Schools, 41 Clusters of Excellence, and ten Institutional Strategies) were selected for the final round. Of these, 38 proposals submitted by 22 universities were selected for funding. They will be funded up to November 2011 at a total of 873 million euro. 305 Draft Proposals were received in the second round of funding, of which 92 (44 Graduate Schools, 40 Clusters of Excellence, eight Institutional Strategies) reached the final round. A total of 47 proposals submitted by 28 universities were selected for funding. They will have received a total of more than one billion euro by November 2012.

Altogether, a total of 1.9 billion euro was granted for the three funding lines in the two rounds in 2006 and 2007. These funds are to be available for the universities and their partner institutions for research and the promotion of young researchers until 2012.

Dissemination, incl. how much information is available regarding data and methods:

A special brochure has been published, presenting all 85 institutions involved in the Initiative for Excellence. Divided into the three competition categories of Graduate Schools, Clusters of Excellence, and Institutional Strategies, and into host institutions, it provides information on the main research topics, on the universities and other institutions participating in the Initiative, along with the names and contact details of the key people involved. The brochure is targeted not only at scientists and researchers, but above all at decision-makers in politics, business, industry and society in general, at media representatives, and last, but not least, at a hopefully large portion of the general public, all of whom are to gain a clear idea of the diversity of topics and issues in, and the viability of university-based research.

Intended and Unintended Consequences:

Observations /Additional Comments:

10.9 HUNGARY

Executive Summary

The research assessment related exercise is based on the Act on Higher Education (2005) and its complementary law (2007). On these bases public HE institutes and the maintainer signed a 3-year contract in 2007 that is to contain a detailed list of measures for monitoring and assessment. The Ministry of Education and Culture as Maintainer specified the obligatory topics for assessment. The HE institutes have to select the relevant indicators by each obligatory field for monitoring and evaluation purposes. After the recognition of systematic weaknesses of the Agreements both Ministry and research-oriented universities wish to modify the indicative indicators for assessment.

Name/Title of Research Assessment Exercise:

Maintainer Agreement between Hungarian Ministry of Education and Culture and HEIs took effect from first of January 2008 and the expiration date is 31st of December 2011. This Agreement specified indicative indicators to assess efficiency of HEI's research activity among other assessment purposes.

Policy Context, incl. circumstances under which the exercise came about:

In the early 1990s, the Hungarian higher education system faced the challenges of double transformation. It had to, on one hand, move away from the (reformed) old socialist system towards the traditional European one and, on the other hand, catch up to the wave of modernisation of European higher education. The adjustment and modernisation of Hungarian HE have been on-going processes since the initial years of transition, which have led to enormous changes both within institutions and in the environment of higher education. Hungary has joined to Bologna process and wishes to become a remarkable actor in the European Higher Education Area.

The general assessment exercise that includes research assessment is based on the Act CXXXIX on Higher Education was passed in 2005 and its complementary law CIV was enacted in 2007. According to these laws a HEI can benefit from public financing if it has a three-year agreement on performing its strategy with the Ministry of Education and Culture.

Hungary has not yet found the right balance between the educational and research autonomy of HE institutions on the one hand and good management of public resources, on the other. The management of HEIs has remained in the hands of academicians. The half-hearted revamping of higher educational governance structures leaves plenty of potential for further improvements and legislation. That is the reason why the Ministry, that has a key role in financing of higher education introduced the Maintainer Agreement system. The Ministry wishes to help the institutes to elaborate and improve their own governing methods and management skills, to strengthen their fact-based strategy making and to promote the activity of HEIs in the European Higher Education Area.

Policy Objective(s):

The main policy objectives are the improvements in quality, competitiveness and efficiency of HEIs. There are two levels of policy objectives:

(1) The Ministry as maintainer can monitor (and assess) capabilities of HEIs for setting and performing strategic targets in various fields during the contracted 3 years. Following this pilot phase Ministry wishes to use the (revised) indicators for comparing HEIs and link stronger the changing part of government subsidy of HEIs to output and outcome indicators. The state subsidy should be fit to the real performance of HEIs.

(2) The Agreement provides a new frame to improve the management and governance of HEIs. With the help of self-selected indicators HEIs can monitor how they are performing their own strategic targets. They can back up put their own strategy-on the base of the facts. By the idea the fact based analysis can significantly improve the management culture and governance of HEIs to coop with changing world.

Methodology, incl. time-frame, resources, costs, technologies:

On the basis of CIV. 2007 (Paragraph 30), public HE institutes and the maintainer signed a 3-year contract at the end of 2007 that is to contain a detailed list of measures for monitoring and assessment. The Ministry of Education and Culture as Maintainer specified the obligatory topics for assessment such as (1) basic activity (Education and Research); (2) supporting activity (guiding and management and collaboration and cooperation); (3) Social linkages (regional role and participation in performing social targets). The HE institutes have to select the relevant indicators by each obligatory field for monitoring and evaluation purposes.

The indicative list for research contains 22 indicators from which each HEI has to select 3-4 indicators by their own strategic plan. Theoretically the institutes could select the best indicators suit to their strategy plan.

Dissemination, incl. how much information is available regarding data and methods:

Beyond the internal dissemination of information has not decided yet. However the aim was making the capabilities and performances of Hungarian HEIs more visible internationally.

Intended and Unintended Consequences:

Good thing is that the institutes somehow are moving toward the fact-based strategy-making and the Ministry also can monitor better the processes and implement the allocation of changing subsidy by norms. However the indicators that are representing less important output activities can slowly support to upgrade the performance of HEIs. The usefulness of selected set of indicators is discussed.

Observations /Additional Comments:

Both, the list of indicative indicators and selected indicators for monitoring and assessment by HEIs are still being debated. In the design stage of the indicators only non-independent experts were involved. In the preparatory phase of Agreement HEIs' selection criteria for indicators was to reach easily good performance if the Ministry is monitoring their activities. The HEIs were much less interested to select indicators for themselves as a managerial tool. Their preference was to choose 'good bargaining' indicator. Actual self-monitoring exercise can hardly support to develop internationally competitive strategy. After the recognition of systematic weaknesses of the Agreements both Ministry and research-oriented universities wish to modify the indicative indicators for assessment. According to the future plans Hungarian HEIs will employ the same research assessment exercise (output and outcome indicators) for both individual and institutional assessment as the institutes of the Hungarian Academy of Sciences (HAS) are employing. The HAS research assessment is an up-to-date version of West European practices.

10.10IRELAND

Executive Summary

The Sunday Times Irish Universities League Table is a relatively basic ranking system, published annually by a private media organisation – it is not a research assessment exercise. Its draws on information provided by HEIs and publicly available data. While its principal aim has been to assist third level entrants and their parents, its influence is beginning to be much wider.

Name/Title of Research Assessment Exercise:

The Sunday Times Irish Universities League Table

http://extras.timesonline.co.uk/tol_gug/gooduniversityguide.php

http://extras.timesonline.co.uk/stug/universityguide.php

Policy Context, incl. circumstances under which the exercise came about:

The Sunday Times is leading newspaper in Britain and Ireland with 1.3m weekly sales. The majority of its readers are graduates. There is some career guidance information provided in the national press and by Guidance Counsellors, and each institution. Otherwise, people relied on local and personal knowledge/intelligence about different institutions, which carried an implicit rank. To fill this gap, *The Sunday Times* six years ago launched the first Irish universities league table in a 94-page University Guide. This also included heavily researched profiles of the main third level institutions and 10 comparison tables. It also provides some comparison of Irish and UK institutions. The principal aim is to assist would-be third level entrants and their parents in making the choice about what college to choose.

Policy Objective(s):

As this is a private media initiative, there is no direct policy objective. The closest parallel has been the Department of Education and Science's refusal to allow performance information about secondary schools to be published. For this 'market', the newspapers have used information about 'feeder schools' as a proxy for performance linked to HE entry.

Methodology, incl. time-frame, resources, costs, technologies:

Universities are ranked according to marks scored in six key performance areas.

Average points required for HE entry. The median Leaving Certificate points obtained by honours degree course entrants, weighted by the latest data on the number of students on each course. A maximum score of 600pts is assumed and the percentage of the maximum attained is given a 2.5 times weighting in the league table. Courses where additional points are awarded for interview/portfolio are excluded. Source: CAO 2008, round 1 data.

Source: Calculated from CAO entry data 2007.

Research. A measure of research efficiency which compares competitive research funding won in 2007 with the number full-time equivalent academic staff. NUI Maynooth had the best ratio, which was scored 100 in the table. All other scores were

then expressed as a percentage of the NUI Maynooth result. Source: individual institutions provided the research income figure for 2007.

Employment. The percentage of graduates known to be seeking employment nine months after graduation. Subtracted from 100 to produce the league table score. Source: Individual colleges extracted from latest available Higher Education Authority (HEA) data.

Firsts/2:1s The percentage of highest quality degrees in 2007. Source: Individual institutions.

Student-staff ratio. Full-time and part-time undergraduate students (weighted), divided by full-time equivalent teaching staff. A ratio of 10:1 as a benchmark for excellence, worthy of 100pts in the league table. Source: Calculated from: Universities HEA 2007 data; institutes 2006 Department of Education and Science data.

Completion rates. Percentage of 2002 entrants who completed courses for which they enrolled by conferring in 2007 Source: Individual Institutes. Trinity College, estimated figure.

OTHER INDICATORS IN THE PROFILES

Undergraduates/postgraduates Full-time undergraduate and postgraduate enrolments. Separate figure in brackets denotes part-time enrolments at each level.

Teaching staff Full-time equivalent number of staff engaged in teaching.

Mature/overseas students Those over 23; those not from Republic of Ireland. Source: HEA 2007 data, universities only.

Non-standard entry Individual institutions 2007 intake.

Sports facilities Assessment by *The Sunday Times* in consultation with students unions. From one star (poor) to five stars (excellent).

Private rents: Student unions â€™ in July 2008.

Time-frame - questionnaires sent to colleges and students' unions in April/May. Guide published in late September.

Resources: One part-time Irish researcher. Two part-time UK ones.

Costs: â,¬7,000 p.a. for Irish research.

Technologies: E-mail, special Excel templates, MS Access database.

Dissemination, incl. how much information is available regarding data and methods:

Results disclosed in the newspaper and on-line.

http://www.timesonline.co.uk/tol/life_and_style/education/sunday_times_university_guide/article2497779.ece

Intended and Unintended Consequences:

Intended consequence: In the absence of other public benchmarking of higher education performance, the *Sunday Times* has filled a gap. As a result, the *Sunday Times* claims that the table is used widely within colleges and by outside partners. Certainly, the publication of the ranking is widely anticipated (and feared) within the higher education community, and it

has introduced a new competitive dynamic into the system despite concern about the indicators and data. No data is available on the extent to which the information is informing student choice.

However, because of the absence of good verifiable and comparable data, the results are controversial. *Sunday Times University Guide* is not a research assessment exercise but has been used as such because it highlights indicators on research.

Unintended consequences: A number of institutions have begun using it as an annual benchmarking exercise and take action to improve their position.

Observations /Additional Comments:

The amount and quality of educational data available in Ireland is poor compared to that in the UK and other countries. In addition to the *Sunday Times* benchmarking of research activity, the results of various national competitions can be interpreted as an assessment of research performance, most notably the Higher Education Authority's Programme for Research in Third-level Institutions (PRTLI) and Science Foundation Ireland (SFI).

The Sunday Times University Guide is not a significant sales driver, typically only adding less than 3% on 1.4m sales. DVDs etc would do much better and be far cheaper to produce. Any increase in sales would be offset by the cost of advertising it and production - 1.3m copies of a 94-page full colour supplement is extremely expensive to produce. It is produced at a loss, or close to it, not to drive short-term sales but to help establish a long-term relationship with 16 to 18 year olds interested in college as future purchasers of the paper. It gets them interested in the type of editorial coverage that the newspaper provides. It is for this reason that *The Irish Times* and *Irish Independent* provide significant information for prospective students on the application process, comparison between programmes, entry level qualifications, etc.

10.11ITALY

Executive Summary

In Italy the evaluation of university-based research witnessed two main periods:

- research produced in the 2001-2003 period has been evaluated by the National Committee for the Evaluation of Research (CIVR, Comitato di Indirizzo per la Valutazione della Ricerca), which published results in 2006;
- in 2009, a new Agency for the Evaluation of University System and Research (ANVUR) has been established by the Government, taking the role of CIVR (evaluation of research) and of CNVSU (National Committee for the Evaluation of University System- evaluation of universities).

Regarding the CIVR experience, the method adopted was self-selection of a sample of scientific products by departments, in proportion to the number of researchers, and single-blind evaluation of products by independent experts. Members of the disciplinary panels established by CIVR used at least 2 independent opinions from international experts, and reached a consensus agreement on a final grade to be assigned to each products. Grades were aggregated at department level and then at university level. The experience was almost unanimously considered positive.

No implication on university funding was derived in subsequent years, until 2009, when the 2001-2003 data were used by the Government to allocate, together with other indicators, 7% of the overall university funding.

Name/Title of Research Assessment Exercise:

Comitato di Indirizzo per la Valutazione della Ricerca (CIVR)

Committee for Steering of Evaluation of Research

Policy Context, incl. circumstances under which the exercise came about:

Established in 1998 (Decree 5 June 1998 No. 204) in order to develop guidelines and implementation of systematic evaluation of research of:

- university departments

Public Research Organisations (CNR, INFN, INFM, ENEA and many others

Policy Objective(s):

To establish common methodologies for evaluation of research (“Guidelines”).
To support internal evaluation bodies at University level and at PRO level to follow common methodologies and to share results.

Carry out systematic and periodic evaluation of research produced by Universities.

Methodology, incl. time-frame, resources, costs, technologies:

CIVR Committee formed by 7 (seven) members.
Staff and collaboration ca. 14 units.
Duration 4 (four) years.
First appointment (Minister Berlinguer) in 1999; second appointment (Minister Moratti) in 2003.
First mandate to evaluate research by Minister of University and Research in 2002.
Guidelines for evaluation of research presented by CIVR in 2003. Approved by Minister of University and Research in December 2003.

Departments were asked to identify a small number of research outputs, to be submitted to the Panels for evaluation. In particular, each University was asked to submit 0,5 products per Full Time Equivalent researcher covering the last three years; each PRO was asked to submit 1 product for FTE. CIVR classified research products in 20 areas, of which 14 disciplinary areas and 6 thematic areas. Each areas was assigned to a Panel, with a Chairman and a number of experts between 5 and 9 units (151 in total), mainly from the national scientific community, supported by 6,661 external experts at international level.

Each product was evaluated by at least two experts. Experts rated products on a four grade scale (Excellent, Good, Acceptable, Unsatisfactory) (the latter being "limited", or "limitato" in Italian language, a rather ambiguous word). Excellent means top 20%; Good 60-80%, Acceptable 40-60%, Unsatisfactory below 40%.
Evaluation was blind (but not double blind, given the published nature of products).
Criteria for evaluation have been the following:

- scientific quality (positioning of the product with respect to scientific excellence)
- relevance (value added of the product for the advancement of knowledge in the field and for science in general; social benefits generated in terms of appropriateness, effectiveness, timeliness and duration of impact)
- originality/ innovativeness
- internationalization or international potential.

Each panellist was responsible for a sub-group of research products. The panellist integrated the two or more external evaluations into a unique score and submitted to the overall panel for the consensus. The evaluation was published only after consensus.

First evaluation was carried out on research products for the period 2001-2003.
All universities and PROs were invited to submit their research products before June 2004, while complementary information on human resources, financial resources, other research outcomes, and evaluation practices were submitted by end of September 2004.
The results were published by CIVR on January 26th, 2006. The Report included the evaluation of each product (n= 18.508), a ranking list of departments, and a methodological report. The ranking list of departments was done by distinguishing departments by size (small, small and medium-sized, large and very large) and reporting the proportion of products in each quality score.

After the last evaluation (2006) there was large expectation that research funding at Ministry level would follow, directly or indirectly, the assessment of CIVR. This expectation did not materialize until 2009 (see below).
Furthermore, no evaluation was demanded by the Ministry for the subsequent periods (e.g. 2004-2006), although the CIVR was ready to take the mandate.
Rather, in 2007 the new Government (Minister of Research Mussi) approved the creation of a new Agency, called National Agency for the Evaluation of the University System and of Research (ANVUR). This new Agency would have cumulated the competences of CIVR for research and of another Committee in charge of evaluating education activities and of supporting the Ministry in funding allocation (CNVSU). Unfortunately, the implementation of the new Agency was difficult, due to several procedural and financial obstacles, so that with the change in Government (April 2008) it was still not operational.

The new Ministry of Research (Gelmini) took office in 2008, declared the need for a streamlining and redesign of the Agency and stopped the implementation. The goal of making research evaluation a pillar of University funding has nevertheless been declared in Government initiatives.

In fact, the Decree no. 180 of November 10th, 2008, stated that a portion of Government funding to universities, not less than 7%, will be done following indicators regarding quality of education and research and efficiency in organization. In addition, it stated that academic staff not producing scientific publications in the last two or three years will not receive full salary upgrade and will exclude from selection boards, respectively.

On 24th July 2009 the Government finally issued a Decree with which 7% of the block funding to universities (FFO, *Fondo di Finanziamento Ordinario*), i.e. 525 million euro, have been allocated to universities according to the following criteria:

- teaching (1/3 of total):

40% number of students at 2nd year with at least 2/3 of CFU in exams of the 1st year;

20% employment of students after 3 years from graduation;

20% use of internal academic staff for teaching

20% adoption of student satisfaction surveys for the evaluation of teaching activities.

- research (2/3 of total):

50% in proportion to the grade received by the University from CIVR in 2006

30% according to share of EU funding (VI and VII Framework programmes)

20% share of funding from Ministry of Research in competitive grant allocations (e.g. PRIN, FIRB).

The Ministry of Research used data from CIVR, CNVSU, and Ministry sources. CIVR data refer to the evaluation of research in the period 2001-2003. The Ministry published a list of universities with the percentage of increase or decrease of funding resulting from the application of these criteria. In the presentation of the criteria, the Ministry declared the intention to increase the proportion of funding allocated according to these criteria.

The same Decree has created the new Agency for the Evaluation of Research (ANVUR), whose President will be nominated by the President of Republic. A list of candidate members will be suggested by a Selection Board of 5 members, nominated by the Ministry of Research, the Secretary General of OECD, the President of *Accademia Italiana dei Lincei*, the European Research Council and the National Council of University Students. Members will be nominated by the President of Republic upon designation of Minister of Research.

Meanwhile, a new Agency for the evaluation of industrial research, not having competence on University-based research, has been implemented in November 2008.

Dissemination, incl. how much information is available regarding data and methods:

Full transparency on methods and mandate to the external experts.
Experts classify research outputs according to classes of quality, departments are ranked according to an aggregation of scores on research outputs.
Universities receive a score representing the proportion of departments ranked top.

Names of experts are kept secret.
CIVR took into account many criticisms and suggestions.

Several meetings and official conferences organized by CIVR together with main PROs, universities and academic communities in order to present methodology and discuss results.

Intended and Unintended Consequences:

The CIVR process had very positive consequences in terms of demonstration of feasibility of the evaluation exercise and credibility of methodology and quantitative scores. In general, few methodological objections were raised with respect to CIVR.
Only in one case (economics) a strong debate was reported within the Evaluation Panel, based on conflict between a view supporting international publications as the exclusive valuable research output, and a view asking for representativeness of national literature and minority (even unorthodox) research traditions. The former view was however dominant.

Among the unintended consequences:

1. failure of the Ministry of Research to identify mechanisms to enforce the evaluation in terms of research funding, leaving some frustration in the academic community supporting the evaluation culture
2. some politics at the level of universities, trying to "immunize" universities from the effect of potential impact of evaluation (i.e. differential funding for good and poor departments) by asking ex ante commitment to redistribute resources equally (but this provision was in fact never implemented).

The initiative of Government in July 2009 was generally welcome, although several contributions in the press stressed that indicators used are questionable (e.g. CIVR evaluation refer to 2001-2003 period) and that universities with very different size, age and subject mix have been treated in the same way.

Regarding the new Agency for Evaluation, it is clearly too early to formulate a judgment.

Observations /Additional Comments:

CIVR extremely positive experience.

Evaluation practice and culture now firmly established in the academic community.

Great expectations on the newly created Agency (2009).

10.12NETHERLANDS

Executive Summary

The three main Netherlands organisations responsible for publicly funded research defined a protocol for practical use in all future research evaluations conducted under their auspices, aimed at assessing all publicly funded research once every six years. A crucial element is that once every three years research units produce a self-evaluation. This case study describes the protocol.

Name/Title of Research Assessment Exercise:

Standard Evaluation Protocol 2003-2009 for public research organisations. A large part of the information provided below is extracted from a document with the same title, published by three Netherlands organisations mentioned below in January 2003.

Policy Context, incl. circumstances under which the exercise came about:

The three main Dutch organisations responsible for publicly funded research – the universities, the Royal Netherlands Academy of Arts and Sciences (KNAW) and the Netherlands Organisation for Scientific Research (NWO) – defined a protocol for practical use in all future research evaluations conducted under their auspices. In this evaluation system all publicly funded research is evaluated once every six years. Once every three years, research units produce a self-evaluation, alternating between preparation for the external evaluation and serving as an internal mid-term evaluation.

The board of the organisation under whose jurisdiction a given research institute falls, i.e. the board of a specific university, of KNAW or of NWO, is responsible for the organisation of the evaluation of that institute and for applying proper procedures.

The evaluation system is a combination of retrospective and prospective analysis. The relationship between retro- and prospective evaluation is to some extent the result of confidence in the future based on insight into the past. In other words: discussions about the future require knowledge of the past. The emphasis is on prospective analysis.

The system aims at operating with the least possible burden for the researchers: a self-evaluation once every three years, an external evaluation once every six years. On the basis of a yearly monitoring system, the institutes maintain data needed for these evaluations in a systematic way. The three research organisations intend to create a national research information system, accessible through the Internet, to store all relevant data.

Policy Objective(s):

The evaluation system aims at three objectives with regard to research and research management:

- Improvement of the quality of research through an assessment carried out according to international standards of quality and relevance;
- Improvement of research management and leadership;
- Accountability to higher levels of the research organisations and funding agencies, as well as to the government, and society at large.

An important condition is also to keep the administrative burden as low as possible. For that reason these evaluations are intended to serve all regular public evaluation goals.

Methodology, incl. time-frame, resources, costs, technologies:

The main criteria to be used in the evaluation are:

Quality (international recognition and innovative potential);

Productivity (scientific output);

Relevance (scientific and socio-economic impact);

Vitality and feasibility (flexibility, management, and leadership).

The evaluation committee presents its judgements on these criteria according to a five-point scale: excellent, very good, good, satisfactory, and unsatisfactory.

Each institute is assessed by an external peer evaluation committee once every six years. The institute produces a self-evaluation every three years, one in preparation of the external review, and he other three years thereafter as a mid-term review. There may be some overlap between different institutes; for example, researchers may work both in an Academy institute and in a university-based research school. It is one of the goals of this system to avoid unnecessary overlap between the evaluations of the various institutes. A leading principle therefore is that information about groups, programmes or parts of the institute evaluated in one evaluation may be used in another. The evaluation committee visits the institute being evaluated. A review is prepared of each research programme run by the institute, containing:

A quantified assessment of the quality, productivity, relevance and prospects of the research programme

An explanation of this quantified assessment, containing:

A reflection on the leadership, strategy and policy of/for the research programme

An assessment of the quality of the research staff, (human) resources, funding policies and facilities

An assessment of the quality and quantity of the publications and of the publication strategies

An assessment of the academic reputation of the group/ programme

An assessment of the relevance of the programme from an academic perspective and from a broader social perspective

An assessment of the future perspectives of the group/ programme.

Data must be provided about funding and resources. The academic reputation of a given institute may be indicated in several ways. Institutes and disciplines may refer to the practice of presenting a bibliometric analysis of the citations of the scientific results. Previous peer reviews, rewards and prizes may also be cited.

In the past few years several evaluation committees, mainly in the natural and life sciences, have used the outcomes of extended bibliometric studies carried out by the Centre for Science and Technology Studies (CWTS).

Dissemination, incl. how much information is available regarding data and methods:

The final evaluation reports are sent to the advisory boards of the institutes evaluated. Each advisory board offers advice on all relevant matters arising from the report received. On the basis of the report, the advisory board's advice and discussions with the institute, the university / KNAW / NWO board will draw conclusions for the future of the institute. The self-evaluation document, the final evaluation report, and the conclusions made by the board together constitute the results of the external evaluation.

The board will report on the results of both the mid-term self-evaluation and the external evaluation in its annual report. The board will make the outcome of the external evaluation available to anyone on request; preferably, results will be made available on the Internet.

Intended and Unintended Consequences:

A systematic account of intended and unintended consequences would require a separate study. To the best of my knowledge such a study has not (yet?) been carried out.

Observations /Additional Comments:

For some disciplines an external peer review was carried out at the national level, evaluating all research groups in all Netherlands universities at the same time (e.g. computer science; chemistry). In other disciplines (e.g. physics), several evaluation committees were established, each dealing only with a limited number of institutions (e.g., Leiden University and Delft University of Technology in the field of physics).

10.13NORWAY

Executive Summary

A new model for result based university research funding was established in Norway in 2006. The main policy objective was to stimulate increased research activities and allocate resources to centers performing excellent research. The methodology is based on three elements: a) base component, b) education component, c) research component. This system will probably result in quite substantial restructuring of the Norwegian research system. An evaluation of the system is foreseen for 2009.

Name/Title of Research Assessment Exercise:

'New Model for Result based University Research Funding'

Policy Context, incl. circumstances under which the exercise came about:

In 2006, Norway adopted a partly new system for allocating funding to universities and colleges. The new model is partly based on science indicators and gives higher weight to research quality and competition than the previous funding system. The purpose of the new model is to stimulate increased research activities by allocating resources to centers that were able to document excellence in research.

Policy Objective(s):

The policy objective is to stimulate increased research activities and research excellence.

Methodology, incl. time-frame, resources, costs, technologies:

The model is based on three components: a) base component, b) education component and c) research component. Part of the research component is result based, where the allocation of funds is based on performance according to an indicator model measuring recorded research results.

The result-based research component consists of four sets of indicators designed to measure research quality: a) Number of doctorates, b) scientific publication 'points', c) funding from Research Council Norway (RCN) and d) funding from the EU's Framework Programme for research.

The result based component has a fixed annual volume. Funds are allocated and reallocated annually, based on the performance of universities and colleges, according to the 4 sets of indicators, during the two previous years. Allocations are made nationally to the universities and colleges, which then decide how to allocate these funds locally.

The indicator for scientific publications was developed by the Norwegian University and College Council, on commission by the Norwegian Government.

Dissemination, incl. how much information is available regarding data and methods:

All information regarding data and methods is available. However, the overall system, as well as its indicator part, is quite complicated and therefore requires quite sophisticated competences on indicators and accounting procedures, including their consequences, to fully understand the workings and consequences of the system, particularly in relation to

strategic decisions at the level of universities and colleges.

Intended and Unintended Consequences:

Intended consequences are, broadly speaking, to improve scientific quality within the Norwegian higher education system. Despite arguments to the contrary, unintended consequences may be incentives to prioritize certain science fields over others because of the differences in the way the design of the GUF model works for different science fields.

The publication indicator has been debated in Norway. An investigation made by a committee (Vagstad Committee) for the Norwegian Ministry of Education (2007) argued that the publication indicator could result in a development away from relatively resource demanding publications to publications requiring relatively less resources. In response to this, the Ministry of Education arguedthat the risks of such tendencies had partly been met by the fact that the publication indicator consisted of two levels. The Ministry is, however, considering an extension of the different kinds of publications that are included in the publication indicator, which for example could result in an inclusion of new textbooks and other kinds of academic production.

The Vagstad Committee also argued that it was '….a weakness of the research component that it concerns all research areas and all institutions. The report indicates that this could result in the favoring of certain research areas and [therefore] suggests an investigation of a segmentation of the research areas in two or three parts [within the system] in the result-based reallocation.' Statistics on higher education seem, however, to indicate that different research areas get relatively similar impact in the research component, which is taken as support for arguing that the indicator set is treating different research areas in a 'fair' way.

A status report on the quality reform was presented in 2008. The report argues that it is still too early to draw any firm conclusions on the results of the new system or on the impacts of the new indicators. An evaluation is foreseen for 2009.

Observations /Additional Comments:

The new Norwegian model for university funding has been one of several inspirations for the new Swedish model for block funding of universities. The two systems accordingly show important similarities, but also potentially important differences. A deeper comparative evaluation of the workings and impacts of the two systems on science excellence and other intended and unintended impacts would be highly interesting and potentially important for research policy within the EU.

10.14SPAIN

Executive Summary

The research outputs of university teachers in terms of publications are evaluated on a voluntary basis every six years. The evaluation criteria differ among scientific fields, and the evaluators take into consideration the quality of publications measured by articles in JCR and their impact factor. In a number of fields, books (Humanities) or patents (Engineering) are also considered. This policy has increased the internationalization of Spanish research publications while probably biasing against research of direct interest to industry or local governments, which is less likely to be published in reputed journals.

Name/Title of Research Assessment Exercise:

Evaluation of Research Activity

Policy Context, incl. circumstances under which the exercise came about:

In 1989, a new scheme was launched to increase the research activities of both university teachers and researchers in public organizations. It formed part of the policy actions undertaken under the auspices of the the 1st National R&D Plan, following the first comprehensive Science Law issued in 1986 (Ley 13, April 14, 1986 for the General Coordination of Scientific and Technical Research).

A mechanism for the evaluation of individual research activities was established, enabling all teachers and researchers fulfilling certain requisites to have their published research results evaluated every six years.

Policy Objective(s):

There was great concern at that time about the quality of university research in a number of areas, and this mechanism was designed to provide an incentive for teachers and researchers to increase the publication of their research results while at the same time reaching worldwide standards. The underlying hypothesis was that this would encourage both research activities per se and the diffusion of results.

Methodology, incl. time-frame, resources, costs, technologies:

Different scientific fields were defined. The breakdown by fields has been modified several times since the initial launch of the exercise. The current breakdown, made in 2006, distinguishes 11 fields: Mathematics and Physics; Chemistry; Cellular and Molecular Biology; Biomedical Sciences; Nature Sciences; Engineering and Architecture; Social, Political and Educational Sciences; Economics and Management Sciences; Law and Jurisprudence; History and Art; Philosophy, Philology and Linguistics.

Every field has slightly different rules regarding the evaluation of research results, and for every field there is a Committee of six to eight experts from different areas within the field. The members of a given Committee, a number of whom are changed every year, analyse the information provided by the researchers and make proposals to the National Evaluation Commission of the Research Activity. Proposals may be positive, which means that the 6-year period (a *sexesium*) is granted, or negative, in which case it is not. The National

Evaluation Commission rarely overturns a proposal made by one of the Committees.

Applicants present, on a voluntary basis, their CVs for the six years for which the evaluation is requested. They must select five publications out of their total, summarize them briefly and provide quality evidence (impact in Journal of Citation Reports, number of citations, etc.). In general terms, the main criterion for a positive evaluation is that a minimum number (2 or 3) of the five publications selected should be JCR publications. In several fields (Humanities), books and book chapters are evaluated on a higher scale, and in Engineering other research results (such as patents) are also considered.

The criteria for evaluating the scientific quality of publications in journals which are not JCR or of publications in other media are all described in the law.

A minimum number of *sexenia* are required for securing tenure or for becoming a member of the commission which grants tenures. *Sexenia* are also very favourably looked upon in the context of applications for research funding under national and regional R&D plans.

Dissemination, incl. how much information is available regarding data and methods:

Because there is not complete transparency in the way each candidate is evaluated by the Committee, there is sometimes an official claim against an evaluation. In this case the Committee (i.e. the Committee nominated for the following year) re-examines the information provided by the person in question and may reconsider the evaluation.

The law which lays down the evaluation process is claimed to be clear and precise; however, the members of the Committees work within wide margins when making their decisions.

The information on the number of *sexenia* every individual has is not publicly available, whereas the aggregate data by faculty, scientific area, etc. are, and are used for comparisons.

Intended and Unintended Consequences:

The process has had very positive consequences with regard to the quantity and quality of research and to the publication of research results in international publications. Therefore, the main objective of the initiative has been attained. Another positive consequence is that several Spanish journals, conscious of the importance of being included in the JCR or other national or international databases, have improved the evaluation and peer review processes.

As an unintended consequence we should highlight the fact that there has been a change in the behaviour of a substantial number of researchers, who now orient their research activities and publication habits mainly towards JCR, the only journals that are considered worthwhile. At the same time, research done directly for industry, or local / regional governments may be suffering because it is not highly valued and, as a result, less likely to be published in renowned international journals.

Observations /Additional Comments:

10.15SWEDEN

Executive Summary

A new model for allocation of university block grants was established in Sweden in the Governmental Research Bill in 2008. The main policy objective was to incentivise strategic university management for increased research quality. The methodology is based on two components: 1) Bibliometric indicators and 2) External funding. It is on the basis of these two components that the increased funding of universities is going to be allocated. This system will probably result in substantial restructuring of the Swedish research system.

Name/Title of Research Assessment Exercise:

'A new model for allocation of resources'

Policy Context, incl. circumstances under which the exercise came about:

Allocation of General University Funds (GUF or Block grants). Established in the Governmental Research Bill October 2008. Based on a proposal (2007) by a Committee investigating the Swedish resource allocation system for research. The system established by the Government is a simpler version of the system proposed by the Committee.

Policy Objective(s):

Quality incentives and strategic independence for higher education institutions.

'The higher education sector has experienced great changes during the past 10-15 years. A challenge is the increased international competition with new requirements on concentration and prioritization in order to keep a high and competitive scientific quality'. 'The resources that are allocated directly to universities and colleges represent the basis for the activities of higher education institutions. These resources should provide the higher education institutions with possibilities to act independently and make strategic decisions concerning long-term directions and priorities of research and research education.' (Res. Bill p.53)

Methodology, incl. time-frame, resources, costs, technologies:

The methodology based on two equally important components: 1) Bibliometric indicators; 2) External research funding. 1) The bibliometric indicator part is based on quite a complex method of assessing scientific performance and comparing it between science fields, using ISI Web of Science, taking into account both (science) 'area adjusted' publication volume and 'field normalized' citations. 2) The external funding part essentially includes all external funding, treating all external funding sources with equal weight. The impact of the model is counter-balanced by giving different weights to different science fields, on the grounds that for one thing they differ in terms of publication and citation propensity and for another they differ in terms of the propensity to attract external funding. In this balancing, Social Sciences and Humanities are given the weight 2.0, Natural Sciences 1.5, Medical and Engineering Sciences 1.0 and other science fields are given the weight 1.1. These counter-balancing coefficients within the model are applied to the whole index, i.e. the total index based on both bibliometric indicators and external funding. The model is based on three-year periods. From 2009 on, additional GUF funds, representing 14 percent of total GUF, will be provided,

which will be allocated in accordance with the new model. From 2010 on, another 10 percent of the present stock of GUF funds will be reallocated on the basis of the new model. This means that from 2010 on close to 25 percent of total GUF will be allocated according to the new competition-based model. It should be noted that in Sweden as much as nearly 50 percent of governmental R&D funding is in the form of GUF. The system is handled by the Ministry of Education with some methodological support from the Swedish Research Council. The administrative costs of the system have not been calculated, but are estimated to be small or limited on the Government side. Costs may be higher, but probably not substantial at the university level, i.e. in addition to the money that would anyway be needed for management purposes.

Dissemination, incl. how much information is available regarding data and methods:

All information about methods is, in principle, available, although they have not yet been published in a way that makes them easily accessible and understandable without advanced technical competence in bibliometric indicator analysis. The bibliometric method used is to be found in an appendix to the above mentioned proposal submitted by a Governmental committee (2007), which recommended the introduction of this kind of a system in Sweden. The economic consequences of adopting the model, for the whole four-year period 2009-2012 for each higher education institution, were already published in the Government bill. However, the detailed descriptions and analyses of the performance of the various universities have not been published and are not yet openly available, although the intention is that both methods and all data should be fully transparent in order to facilitate improvements in university research.

Intended and Unintended Consequences:

Intended consequences are to generate incentives and resources for universities to prioritize, manage and perform research in a way that improves the scientific quality and, as a consequence, the attractiveness of Swedish research environments in terms of external research funding, research cooperation and talented researcher and student inflow. Intended consequences are also to improve the utilization of research in society and industry. It is probably also intended, although not explicitly stated, that the incentives and resources generated will initiate processes leading to a restructuring of the Swedish higher education system (including possible mergers or alliances between higher education institutions), driven by targets related to scientific quality.

Unintended consequences may be incentives to prioritize certain science fields over others because of the differences in the way the design of the GUF model works for different science fields. Particularly critical in this is the different weights given to different science fields, in combination with the bibliometric indicators and the external funding accounting. The different weights given to different science fields were not motivated by a highly sophisticated and solid rationale, which would make it obvious that these were the ‘right’ or ‘fair’ weights given at the time the system was being put into effect, and this will be even less certain in the future, when global science patterns may change. And as the system is intended to change the behaviour of Swedish higher education institutions, their allocations between science fields and the publication and external orientation within science fields may change as a consequence of the differences in terms of the ways the model rewards performance within different fields. These changes may alter the justifications for the

different weights given to different science fields.

Observations /Additional Comments:

The new model for allocating GUF in the Swedish higher education system can be expected to generate strong restructuring impacts on the system. The directions, patterns and magnitudes of such restructuring processes are, however, difficult to predict. It is also difficult to foresee how efficient this model will be in improving the international competitiveness of Swedish research in terms of scientific excellence or in terms of its benefits for society and industry.

10.16UNITED KINGDOM

Executive Summary

Since 1986 the UK national funding bodies have evaluated the quality of research in UK universities through peer review and have used the results to inform the selective distribution of public funds. The most recent Research Assessment Exercise (RAE) took place in 2008 and will inform research funding for universities in 2009-10. Inevitably, the RAE results are converted by the media into league tables for ranking the quality of subject areas and universities. The RAE is developing into a new Research Excellence Framework which has the intention of blending a lighter touch peer review with bibliometric indicators where these are appropriate.

Name/Title of Research Assessment Exercise:

Research Assessment Exercise 2008

Policy Context, incl. circumstances under which the exercise came about:

The first RAE was undertaken in 1986 and introduced an explicit and formalised assessment process of the quality of research. Further exercises held in 1989, 1992 and 1996 became gradually more transparent, comprehensive and systematic. The fifth exercise in 2001 considered the work of almost 50,000 researchers in 2,598 submissions from 173 higher education institutions. The RAE is the principal means by which institutions assure themselves of the quality of research undertaken in the HE sector. The current research assessment exercise, RAE 2008, published the results on 18 December 2008.The RAE is a peer review exercise to evaluate the quality of research in UK higher education institutions. This assessment informs the selective distribution of funds by the UK higher education funding bodies.

Policy Objective(s):

The primary purpose of the RAE 2008 is to produce quality profiles for each submission of research activity made by higher education institutions. The four higher education funding bodies in the UK intend to use the quality profiles to determine the grant for research for each higher education institution with effect from 2009-10.

Methodology, incl. time-frame, resources, costs, technologies:

RAE 2008 - Peer review with panels meeting to determine a collective view on the quality of research for each submitted unit of assessment in each higher education institution. The main body of assessment took place in 2007-08. A central RAE team based within the HEFCE manages the process for all four UK funding bodies.

Dissemination, incl. how much information is available regarding data and methods:

The RAE 2008 assessment method is transparent and all aspects of the methodology are in the public domain. Full information about RAE 2008 can be obtained from the HEFCE website – www.hefce.ac.uk -The RAE results are made available publicly both by the funding bodies and the media.

Intended and Unintended Consequences:

It is not the intention of the funding bodies to create league tables from the RAE 2008 results but that is the likely consequence through the media. The intended consequence for the funding bodies is to find a way of assessing the quality of research and then linking that quality judgement to funding in a way that commands the confidence of the higher education sector.

Observations /Additional Comments:

The Higher Education Funding Council for England is working to develop new arrangements for the assessment and funding of research. The new arrangements – the Research Excellence Framework (REF) – will be introduced after the RAE 2008. The REF will consist of a single unified framework for the funding and assessment of research across all subjects. It will make greater use of quantitative indicators in the assessment of research quality than the present RAE system, while taking account of key differences between the different disciplines. Assessment will combine quantitative indicators – including bibliometric indicators wherever these are appropriate – and light touch expert review. Which of these elements are employed, and the balance between them, will vary as appropriate to each subject. The REF will be developed as a single unified framework throughout 2009. Aspects of the framework will be phased in from 2011-12, and it will fully drive research funding for all disciplines from 2014.

10.17GLOBAL - WEBOMETRICS

Executive Summary

Webometrics, an initiative of the Cybermetrics Lab (Spain), has produced the 'World Universities' ranking on the Web' since 2004, measuring the web presence of universities around the world and comparing the size and scale of their web presence against expectations based on other rankings. The Webometrics ranking is now updated every 6 months, with data collection occurring in January and July, and the results of the data analysis are published a month later. The indicators correlate web measures with traditional scientometric and bibliometric indicators used in other rankings.

Name/Title of Research Assessment Exercise:

Webometrics, CINDOC, Spain

Policy Context, incl. circumstances under which the exercise came about:

History:

Webometrics produces the 'World Universities' ranking on the Web,' which calculates the web presence of universities around the world and compares the size and scale of the web presence against expectations based on other rankings. 'Web presence measures the activity and visibility of the institutions and it is a good indicator of impact and prestige of universities' (Webometrics, 2008).

The Webometrics world universities ranking are initiatives of the Cybermetrics Lab, which is a research group within the Centro de Información y Documentación (CINDOC). CINDOC is a subsidiary body within the National Research Council (CSIC), the largest public research body in Spain.

Officially launched in 2004, the Webometrics ranking is now updated every 6 months, with data collection occurring in January and July, and the results of the data analysis are published a month later. The indicators used correlate web indicators with traditional scientometric and bibliometric indicators. The goal of the webometrics project is to showcase the importance of the web for the academic community, for dissemination of academic knowledge as well as for measuring scientific activities, performance, and impact.

Policy Objective(s):

Objectives:

Initially, Webometric's analysis aimed to simply highlight the significance of Web presence and publication. CINDOC supports Open Access initiatives, promoting electronic access to scientific publications and to other academic material. Using this data for a ranking was something of an afterthought, but ultimately made sense, as the web indicators used in the analysis are not based on number of page hits or site design but instead on universities' web presence as illustrative of institutional outputs and web visibility.

'We intend to motivate both institutions and scholars to have a web presence that reflect accurately their activities. If the web performance of an institution is below the expected position according to their academic excellence, university authorities should reconsider their web policy, promoting substantial increases of the volume and quality of their

electronic publications.'

'However web indicators are very useful for ranking purposes too as they are not based on number of visits or page design but global performance and visibility of the universities.

'As other rankings focused only on a few relevant aspects, specially research results, web indicators based ranking reflects better the whole picture, as many other activities of professors and researchers are showed by their web presence.

The Web covers not only formal (e-journals, repositories) but also informal scholarly communication. Web publication is cheaper, maintaining the high standards of quality of peer review processes. It could also reach much larger potential audiences, offering access to scientific knowledge to researchers and institutions located in developing countries and also to third parties (economic, industrial, political or cultural stakeholders) in their own community.

The Webometrics ranking has a larger coverage than other similar rankings (see table below). The ranking is not only focused on research results but also in other indicators which may reflect better the global quality of the scholar and research institutions worldwide.

We intend to motivate both institutions and scholars to have a web presence that reflect accurately their activities. If the web performance of an institution is below the expected position according to their academic excellence, university authorities should reconsider their web policy, promoting substantial increases of the volume and quality of their electronic publications.'

'The project intends to have true global coverage, not narrowing the analysis to a few hundreds of institutions (world-class universities) but including as many organizations as possible. The only requirement in our international rankings is having an autonomous web presence with an independent web domain.'

'With these rankings we intend to provide extra motivation to researchers worldwide for publishing more and better scientific content on the Web, making it available to colleagues and people wherever they are located.'

'…the current objective of the Webometrics Ranking is to promote Web publication by universities, evaluating the commitment to the electronic distribution of these organizations and to fight a very concerning academic digital divide which is evident even among world universities from developed countries. However, even when we do not intend to assess universities performance solely on the basis of their web output, Webometrics Ranking is measuring a wider range of activities than the current generation of bibliometric indicators that focuses only in the activities of scientific elite.'

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

The success of self-archiving and other repositories related initiatives can be roughly represented from rich file and Scholar data. The huge numbers involved with the PDF and doc formats means that not only administrative reports and bureaucratic forms are involved. PostScript and Powerpoint files are clearly related to academic activities.'

Methodology, incl. time-frame, resources, costs, technologies:

‘This ranking has the largest coverage with more than 16,000 Higher Education Institutions worldwide listed in the Directory.’

‘Webometrics Ranking is measuring the volume, visibility and impact of the web pages published by universities, with special emphasis in the scientific output (referred papers, conference contributions, pre-prints, monographs, thesis, reports, …) but also taking into account other materials (courseware, seminars or workshops documentation, digital libraries, databases, multimedia, personal pages, …) and the general information on the institution, their departments, research groups or supporting services and people working or attending courses.’

‘Access to the Web information is done mainly through search engines. These intermediaries are free, universal, and very powerful even when considering their shortcomings (coverage limitations and biases, lack of transparency, commercial secrets and strategies, irregular behaviour). Search engines are important for measuring visibility and impact of university’s websites.

There are a limited number of sources that can be useful for Webometric purposes: 7 general search engines (Google*, Yahoo Search*, Live (MSN) Search*, Exalead*, Ask (Teoma), Gigablast and Alexa) and 2 specialised scientific databases (Google Scholar* and Live Academic). All of them have very large (huge) independent databases, but due to the availability of their data collection procedures (Apis), only those marked with asterisk are used in compiling the Webometrics Ranking.’

‘The unit for analysis is the institutional domain, so only universities and research centres with an independent web domain are considered. If an institution has more than one main domain, two or more entries are used with the different addresses. About 5-10% of the institutions have no independent web presence, most of them located in developing countries.’

‘So the best way to build the ranking is combining a group of indicators that measures these different aspects. Almind and Ingwersen proposed the first Web indicator, Web Impact Factor (WIF), based on link analysis that combines the number of external inlinks and the number of pages of the website, a ratio of 1:1 between visibility and size. This ratio is used for the ranking but adding two new indicators to the size component: Number of documents, measured from the number of rich files in a web domain, and number of publications being collected by Google Scholar database. As it has been already commented, the four indicators were obtained from the quantitative results provided by the main search engines as follows:

Size (S). Number of pages recovered from four engines: Google, Yahoo, Live Search and Exalead. For each engine, results are log-normalised to 1 for the highest value. Then for each domain, maximum and minimum results are excluded and every institution is assigned a rank according to the combined sum.

Visibility (V). The total number of unique external links received (inlinks) by a site can be only confidently obtained from Yahoo Search, Live Search and Exalead. For each engine, results are log-normalised to 1 for the highest value and then combined to generate the rank.

Rich Files (R). After evaluation of their relevance to academic and publication activities and considering the volume of the different file formats, the following were selected: Adobe

Acrobat (.pdf), Adobe PostScript (.ps), Microsoft Word (.doc) and Microsoft Powerpoint (.ppt). These data were extracted using Google and merging the results for each filetype after log-normalising in the same way as described before.

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

The four ranks were combined according to a formula where each one has a different weight:

Webometrics Rank (position)=

4*RankV+2*RankS+1*RankR+1*RankSc'

Dissemination, incl. how much information is available regarding data and methods:

Entirely web disseminated, with data and methodology clearly presented and articulated on their website.

Intended and Unintended Consequences:

Limitations:

'The use of link farms and paid backlinks to improve the position in our Webometrics Rankings is not acceptable as this is a non academic practice and it is contrary to the aims of this Ranking. The involved institutions do not have a place in our Ranking and will not be classified in future editions. Random checks are made to ensure the correctness of the data obtained.'

'Current identified biases of the Webometrics Ranking includes the traditional linguistic one (more than half of the internet users are English-speaking people), and a new disciplinary one (technology instead of biomedicine is at the moment the hot topic) Since in most cases the infrastructure (web space) and the connectivity to the Internet already exists, the economic factor is not considered a major limitation (at least for the 3,000 Top universities).'

'The only source for the data of the Webometrics Ranking is a small set of globally available, free access search engines. All the results can be duplicated according to the describing methodologies taking into account the explosive growth of the web contents, their volatility and the irregular behaviour of the commercial engines.'

Observations /Additional Comments:

This information is:

of a general nature only and is not intended to address the specific circumstances of any particular individual or entity;

not necessarily comprehensive, complete, accurate or up to date;

sometimes linked to external sites over which consortium members have no control and for which they assume no responsibility;

not professional or legal advice (if you need specific advice, you should always consult a suitably qualified professional).

http://www.webometrics.info/disclaimer.html

10.18GLOBAL - ACADEMIC RANKING OF WORLD UNIVERSITIES (ARWU)

Executive Summary

The Academic Ranking of World Universities (ARWU), first published in 2003 and updated annually by the Institute of Education at Shanghai Jiao Tong University (China), was the first international league table, offering a scientific mechanism for comparing universities around the world. ARWU uses several comparative and seemingly objective indicators of academic or research performance, including alumni and staff winning Nobel prizes and fields medals, highly cited researchers, articles published in Nature and Science, articles indexed in major citation indices, and per capita academic performance of an institution.

Name/Title of Research Assessment Exercise:

Shanghai Jiao Tong University (SJTU)— Academic Ranking of World Universities (ARWU)

Policy Context, incl. circumstances under which the exercise came about:

History:

The Academic Ranking of World Universities (ARWU) was the first internationally comparative league table, developed to offer a scientific mechanism for comparing universities around the world. Starting in 2001, researchers in the Institute of Higher Education at Shanghai Jiao Tong University (SJTU) were compelled to develop a quantitative mechanism for comparing Chinese higher education, '…to assess the gap between Chinese universities and world class universities…' (Liu and Cheng, 2005, 127). As part of a comprehensive public agenda for development within in higher education sector, China was committed to promoting world-class standards among at least its best institutions.

In order to know whom to utilize as comparator institutions, and to know how far China needed to go in investment and development, the SJTU researchers developed ARWU. As the SJTU researchers sought information and advice about building their comparator model, they received requests to share their findings. Those findings, then, became the first public launch of the ARWU, in August 2003. The ranking has been updated each August since, with the latest update, featuring the review of more than 2,000 institutions, published in August 2008.

Policy Objective(s):

According to SJTU's Academic Ranking of World Universities website, the objective of their ranking is to fill a gap in the global information on higher education. Given the expansive range of stakeholders in higher education, including students and their families, academic staff, institutions, governments, and the taxpaying public that has a form of investment in their government-funded higher education opportunities; the challenge of developing a ranking to suit such disparate interests was acute. Each of these groups, as well as others not listed, are interested in rankings of universities for different purposes.

At the time of its inception as an idea in 2001, and remaining true through its first publication in 2003, there was no ranking of world universities using multiple criteria to establish a single perspective on research output and institutional excellence—a focus on the academic environment of an institution—across nations and national systems. The

objective, then, of the ARWU was very purposefully to focus attention on the academic capacity of institutions and on those institutions doing it better than anyone else.

Methodology, incl. time-frame, resources, costs, technologies:

The methodology used by Shanghai Jiao Tong University to develop its own World University Rankings relies exclusively on seemingly objective indicators, such as the academic and research performance of faculty, alumni, and staff. The measures evaluated include publications, citations, and international awards, such as Nobel prizes and field medals.

"We have scanned every university that has any Nobel Laureates, Highly Cited Researchers, or papers published in Nature or Science. In addition, we scanned major universities of every country with significant amount of papers cited by SCIE and SSCI. In total, we have collected data on about 2000 universities." (http://ed.sjtu.edu.cn/rank/2003/FAQ.htm retrieved November 5, 2008.)

Universities are ranked by several indicators of academic or research performance, including alumni and staff winning Nobel prizes and field medals, highly cited researchers, articles published in Nature and Science, articles indexed in major citation indices, and the per capita academic performance of an institution. Attempt to minimize subjectivity of reputation rankings by focusing on output.

Table 1. SJTU-ARWU Methodology

Criteria	Indicator	Code	Weight
Quality of Education	Alumni of an institution winning Nobel prizes and field medals	Alumni	10%
Quality of Faculty	Staff of an institution winning Nobel prizes and field medals	Award	20%
	Highly cited researchers in 21 broad subject categories	HiCi	20%
Research Output	Articles published in Nature and Science*	NandS	20%
	Articles indexed in Science Citation Index-expanded, and Social Science Citation Index Performance	PUB	20%
Per Capita	Per capita academic performance of an institution	PCP	10%
Total			100%

Source: http://www.arwu.org/rank2008/ARWU2008Methodology(EN).htm, retrieved 7 November 2008.

Dissemination, incl. how much information is available regarding data and methods:

Dissemination is mainly through its web site, though the annual release of the ARWU garners extensive publicity in newspapers and journals world-wide. The website contains clear links and descriptions of data and methodology used. The actual data analyzed are not made available, however.

Intended and Unintended Consequences:

Its developers define the ARWU as an academic ranking and not a comprehensive one. ARWU does not explore the overall operational capacity of universities with regard to such inputs admission rates or the educational backgrounds of its faculty. Unlike other rankings, the ARWU omits subjective inputs like reputational factors. Instead, the ARWU ranking focuses specifically on the academic outputs (scholarly works and awards for high quality scholarship and research).

The most obvious bias in this ranking is toward research outputs and not teaching quality and student outputs. This bias is made clear in the methodology, however, and the limitations to research output and the scholarly capacity of academic staff works toward giving ARWU a relatively solid level of credibility among a broad array of higher education stakeholders, in particular, researchers and policymakers. This limitation to more objective data is what also gives this SJTU-ARWU ranking its strength and reputation as the most reliable among the global rankings.

Ultimately, however, the quality of universities cannot be measured precisely using imprecise, basic numbers. The researchers at SJTU are clear in disclaimers on their website that it would be impossible to have a comprehensive ranking of universities worldwide, because of the huge differences of universities, in the large variety of countries and funding capacities, and the technical difficulties in obtaining internationally comparable data.

According to the SJTU ARWU website, 'People should be cautious about any ranking including our Academic Ranking of World Universities. Nevertheless, our Academic Ranking is based on internationally comparable data that everyone could check.' A 2007 article by Rãzvan V. Florian, in Scientometrics, found, in fact, that the results emerging from the ARWU data were not replicable, calling into question the comparability and methodology of the data used in the ranking.

One final bias that deserves mention is that related to the use of English as the language of international scholarship. As citations depend on having access to published scholarship, and the preponderance of published scholarship occurs in English, an unfortunate outcome of this use of citation is the bias against institutions which operate outside English-language countries or areas.

Observations /Additional Comments:

Liu, Nian Cai and Cheng, Ying (2005) 'The Academic Ranking of World Universities', *Higher Education in Europe*, vol. 30, no. 2, pp. 127-135.

10.19GLOBAL - THE-QS WORLD UNIVERSITIES RANKING

Executive Summary

In November 2004, the Times Higher Education Supplement published its first *World University Ranking (WUR)*, a listing of the top 200 institutions across the globe. Using subjective inputs—peer reviews from academics and employers—and quantitative data, such as the numbers of international students and faculty, and the influence of the faculty, as represented by research citations, to compare the international stature of institutions, the *WUR* intends to meet the needs of consumers (students, academic staff, researchers, policy makers) seeking reliable information about universities around the world.

Name/Title of Research Assessment Exercise:

THE-QS World Universities Ranking

Policy Context, incl. circumstances under which the exercise came about:

History:

The Times (London) and its allied publication *The Times Higher Education Supplement (THES)* (now just the *Times Higher Education, THE*) developed basic subject-based reviews of UK institutions during the 1980s. These reviews evolved into data driven, comprehensive national institutional rankings (*Times Good University Guide*) in the 1990s.

In November 2004, the Times Higher published its first *World University Ranking (WUR)*, a listing of the top 200 institutions across the globe, as well as smaller regional rankings. *THE* partnered with a British research firm, QS, to produce the data used in the rankings. Together, QS and *THE* have published the *WUR* for 5 years, the most recent being autumn 2008, with modifications to the methodology each year in a continued effort to improve.

Policy Objective(s):

The stated goal of these rankings was to meet the needs of consumers of their publications and of new consumers who were seeking reliable information about universities across the country. The expansion from producing national rankings to developing an international one resulted from the recognition that student mobility was on the rise at a time when government funding for HE was diminishing. The competition for international students was going to surge as a result, and these students would need a mechanism to inform their choices for international study. The *WUR* was developed specifically to fill this niche consumer market.

Methodology, incl. time-frame, resources, costs, technologies:

The methodology for this ranking supports a more holistic and subjective view of universities' relative strengths by seeking comparisons based on international reputation. The major areas of analysis are teaching, research, and global presence, with the quality of each determined by a combination of qualitative, subjective inputs—peer reviews from academics and employers—and quantitative data, such as the numbers of international students and faculty and the influence of the faculty, as represented by research citations, to compare the international stature of institutions. (See Table 1.)

Table 1.

Indicator	Explanation	Weighting
Academic Peer Review	Composite score drawn from peer review survey (which is divided into five subject areas). 6,354 responses in 2008.	40%
Employer Review	Score based on responses to employer survey. 2,339 responses in 2008.	10%
Faculty Student Ratio	Score based on student faculty ratio	20%
Citations per Faculty	Score based on research performance factored against the size of the research body	20%
International Faculty	Score based on proportion of international faculty	5%
International Students	Score based on proportion of international students	5%

Source:
http://www.topuniversities.com/worlduniversityrankings/methodology/simple_overview/
retrieved 6 November 2008.

Dissemination, incl. how much information is available regarding data and methods:

Annually, this ranking is disseminated in the following ways: in the print version of THE, which is among the best selling editions of THE in any given year, on the THE website, and on the QS website. The information about its data used and methodology is on the website. The actual data analyzed are not made available, however.

Intended and Unintended Consequences:

The *WUR*'s limitations lie in the same breadth of data that the QS/THE developers cite as its strengths—the inconsistency and variability of its findings year to year. Over the five years of production, QS/THE have sought methods to tighten and strengthen its analysis. Over the past few years, however, a disproportionate number of institutions from within the UK have risen to the top of the table, giving a perception of bias in the methodology.

It is also fundamentally biased toward historically significant institutions, since reputation is so heavily weighted (see Table 1). It will remain daunting for newly developing institutions to crack into the highest levels of this ranking, as their reputational factors would need to overtake those of more established institutions.

And, finally, the commercial nature of the *WUR*, with consumers required to buy the paper to access the data, at worst calls the integrity of the process into question, as anticipating consumer demands might in so way influence the methodological choices made. As the methodology has changed annually, this potential conflict of interest seems as valid today as

a few years ago, at the time these rankings first really appeared.
Observations /Additional Comments:

10.20 GLOBAL - PERFORMANCE RANKING OF SCIENTIFIC PAPERS FOR RESEARCH UNIVERSITIES

Executive Summary

First published online in 2007, the Performance Ranking of Scientific Papers for Research Universities (PRSP) was developed by the Higher Education Evaluation and Accreditation Council of Taiwan to gauge the research productivity of the best universities in the world. PRSP employs bibliometric methods to analyze and rank the scientific papers performance of the top 500 universities in the world from an overall listing of 3000 institutions. The performance measures are composed of numerous indicators within three different criteria of scientific papers' performance: research productivity, research impact, and research excellence.

Name/Title of Research Assessment Exercise:

Performance Ranking of Scientific Papers for Research Universities

Higher Education Evaluation and Accreditation Council of Taiwan

Policy Context, incl. circumstances under which the exercise came about:

History:

First published online in 2007 and now in its second year, the Performance Ranking of Scientific Papers for Research Universities (PRSP) was developed to gauge the research productivity of the best universities in the world. As opposed to other more comprehensive rankings, PRSP does not seek to establish a hierarchy of diminishing excellence, with the top being the best and the bottom the worst. Instead, PRSP tracks the academic outputs to provide some comparative data on the work produced by institutions and its utility to the community outside its campus.

Policy Objective(s):

'This performance ranking targeted research-oriented universities, especially those in newly developed countries. Through objective indicators which would also reflect short-term efforts, each university would be able to understand its position and advantages in the world rankings, and from there it would know how it fares against other universities, and it can track its annual progress in terms of the quality and quantity of its scientific papers' (HEEACT, 2007).

Based on objective data obtained to measure both the qualitative and quantitative impact scientific papers, PRSP then utilizes quantitative analytical indicators to illustrate objective characteristics. Once these quantitative data are generated, the PRSP staff use that data in conjunction with concepts that would observe the quality of the papers, making the PRSP a ranking that creates a quantitative deductive ranking utilizing qualitative assessments.

Methodology, incl. time-frame, resources, costs, technologies:

PRSP employs bibliometric methods 'to analyze and rank the scientific papers performances of the top 500 universities in the world' (HEEACT, 2008). From an initial listing of 3000 institutions, the PRSP then uses the number of published articles per institution, as gleaned

from the Essential Science Indicators (ESI), to reduce the list further. The list then is reduced to exclude non-university institutions, and the remaining institutions are compared against other rankings, resulting in a final listing of approximately the top 500 institutions. PRSP, finally, compares these universities' outputs using data from ISI's ESI, *Web of Science* (WOS), which includes SCI and SSCI, and *Journal Citation Reports* (JCR).

> The features of this performance ranking are that 80% are qualitative indicators (the first time using h-index to reflect on the quality of universities' papers), and that long-term and short-tem indicators each (have) weights up to 50%, thereby emphasizing the outcome of short-term efforts. However, this performance ranking focused on the qualitative and quantitative performance of scientific papers, and did not include other frequently used university evaluation indices such as teaching, research, and administration, nor did it emphasize on academic performance indices such as reputation and extraordinary achievements. The indices designed for this performance ranking study were suitable as reference especially for research-oriented universities in newly developed countries. HEEACT, 2008

The performance measures are composed of nine indicators representing three different criteria of scientific papers performance: research productivity (the number of articles published in peer-reviewed academic journals), research impact (the number of citations on a particular academic article within a specific time frame), and research excellence (utilizing the h-index of the last two years, the number of Highly Cited Papers from ESI, the number of articles in high-impact journals in the current year (Hi-Impact journal articles), and the number of subject fields where the university demonstrates excellence(fields of excellence).

The PRSP staff analyzes all the SCI/SSCI bibliographic records in which the address field contained one of the known forms of the university name, removing duplicate records containing different forms of that university's name. They obtain the total number of citations by adding the number of citations on each of the articles from that university from its inclusion in SCI/SSCI to date.

To determine which citations are from separate, independent campuses within a single university system, which are often combined into one SCI/SSCI listing, PRSP staff researchers identify manually the actual number of articles and citations within the SCI/SSCI to identify those produced by each individual campus. This staff-intensive approach is also used to ensure the 'highly cited ESI' calculations are representative of overall institutional quality and not an outlier.

Dissemination, incl. how much information is available regarding data and methods:

Dissemination is through the website, and the data and methods used in this ranking are explained there. Specific data are not presented, however.

Intended and Unintended Consequences:

Limitations:

When universities obtain similar scores, the slight differences of the final scores may not necessarily suggest its superiority in scientific research. The hierarchical nature of the performance ranking may insinuate greater distinction in quality comparators than are

accurate.

The PRSP also does not seek to evaluate the outputs in the humanities and social science, and focuses specifically on performance in the hard sciences. This focus on hard science also skews the ranking in favour of large institutions, for the sheer capacity in funding and faculty to publish, and institutions with medical schools, which publish much higher numbers and with more immediate applicability (citation potential) than many other scientific fields.

And, finally, an important acknowledgement from the PRSP:

> This performance ranking is neither a 'reputation ranking' nor an 'academic ranking' of universities, so some people might feel that the outcome might be different from what they had come to know and expect. In addition, for those universities in a given country that were included in the ranking, there might have been some discrepancies in their actual rank and what was expected of them, but the relative position in rank between universities of the same country may still match society's expectations.
>
> HEEACT, 2007

Observations /Additional Comments:

http://210.71.47.3/ranking/EngMethod.htm

http://ranking.heeact.edu.tw/en-us/2008/Page/Methodology

10.21GLOBAL – THE LEIDEN RANKING

Executive Summary

The Centre for Science and Technology Studies (CWTS) at Leiden University has developed a ranking system of universities entirely based on its own bibliometric indicators. The work focuses on all universities worldwide with more than 700 Web of Science indexed publications per year. About 1000 largest (in terms of number of publications) universities in the world are covered.

Name/Title of Research Assessment Exercise:

The Leiden Ranking

Policy Context, incl. circumstances under which the exercise came about:

During the past years, rankings of world universities have become increasingly important and popular. At a global level universities increasingly compete to attract the best students and research workers, and the European commission launched the concept of a European Research Area. The press started publishing rankings of higher education institutions in order to enable potential clients or customers to make informed choices. The Centre for Science and Technology Studies (CWTS) at Leiden university is specialised in the development and application of research assessment methodologies based on bibliometric indicators. Its members developed a critique on existing rankings of universities such as the Shanghai (ARWU) Ranking and the QS World Universities Ranking, especially on their bibliometric components. They applied the methodologies they had developed in many bibliometric studies of individual universities during the past two decades to a large collection of world universities.

Policy Objective(s):

CWTS aimed at producing rankings that, based solely on bibliometric indicators, reflect a university's recent research performance more adequately than the other ranking systems do. A second objective of the CWTS ranking system was to show how rankings depend upon the type of indicator used, and upon the criteria applied to select universities in the ranking.

Methodology, incl. time-frame, resources, costs, technologies:

The ranking list proposes four different indicators, each resulting in a different ranking as follows:

Total number of publications;

Number of citations per publication;

The number of citations per publication divided by the average impact in the given field. This is the 'crown indicator' because standardisation using the impact of publications in the given field prevents non-generalist universities such as engineering schools or technical universities from being penalised by the citations-per-publication calculation.

Total number of publications multiplied by the relative impact in the given field

The system presents these for rankings for collections of universities:

Europe: the 100 and the 250 largest universities in Europe for the period 2000-2007,

World: the 100 and the 250 largest universities worldwide for the period 2003-2007,

These two different size thresholds illustrate clearly how smaller universities that are not present in the top-100 (in size) may take high position in impact ranking if the size threshold is lowered.

Bibliometric data are extracted from a bibliometric version of Thomson Reuters' Web of Science, created at CWTS. CWTS aims at updating its ranking system every year.

Dissemination, incl. how much information is available regarding data and methods:

The CWTS ranking system is publicly available through the following website: http://www.cwts.nl/ranking/LeidenRankingWebSite.html.

Intended and Unintended Consequences:

The actual use and the effects of the Leiden Ranking have not yet been systematically analyzed.

Observations /Additional Comments:

More information about the methodology underlying the Leiden ranking System and a first analysis of the results are presented in the following research publication:

Calero-Medina C., López-Illescas C., Visser M.S., Moed, H.F. (2008). Important factors in the interpretation of bibliometric rankings of world universities. *Research Evaluation 17*, 71-81.

11 Appendix V. Bibliography

This list of publications does not represent a comprehensive bibliography. Not all publications listed below directly referenced in the report. Rather, the Bibliography is intended to provide some guidance as to material referenced in the report or to other material which might be useful.

EU Publications

European Council (2000) *Presidency Conclusions*, Lisbon 23-24 March 2000, retrieved 28 July 2009, from http://www.consilium.europa.eu/ueDocs/cms_Data/docs/pressData/en/ec/00100-r1.en0.htm

European Council (2005) *Presidency Conclusions*, Brussels. 7619/1/05. Rev 1. Retrieved 28 July 2009, from http://www.clubofrome.at/events/2006/brussels/files/lisbon-strategy-2005.pdf.

Council of the European Union (2007) *Council Resolution on modernising universities for Europe's competitiveness in a global knowledge economy*, 16096/1/07 REV 1. Retrieved 26 July 2009, from http://www.consilium.europa.eu/ueDocs/cms_Data/docs/pressData/en/intm/97237.pdf

CREST 4th OMC Working Group (2009) 'Mutual learning on approaches to improve the excellence of research in universities' Executive Summary.

European Commission (2005a) *European Universities. Enhancing Europe's Research Base*, DG Research, Brussels. Report by the Forum on University-based Research. Retrieved 26 July 2009, from http://ec.europa.eu/research/conferences/2004/univ/pdf/enhancing_europeresearchbase_en.pdf.

European Commission (2005b) *Working together for growth and jobs: A new Start for the Lisbon Strategy*. Communication from President Baroso in agreement with Vice-President Verheugen to the Spring European Council., COM(2005) 24 final. Retrieved 26 July 2009, from http://www.clubofrome.at/events/2006/brussels/files/lisbon-strategy-2005-cc.pdf.

European Commission (2005c) *Implementing the Community Lisbon Programme. More Research and Innovation – Investing for Growth and Employment; A Common Approach, COM(2005) 488 final. Retrieved http://eur-lex.europa.eu/LexUriServ/site/en/com/2005/com2005_0488en01.pdf*

European Commission (2006) *Delivering on the modernisation agenda for universities: Education, research and* innovation. Communication to the Council and the European Parliament, (COM) 2006 208 final. Retrieved 11 August 2009, from http://www.madrimasd.org/proyectoseuropeos/documentos/doc/education_research_and_innovation.pdf.

European Commission (2006) *Encourage corporate measuring reporting on research and other forms of intellectual capital.* RICARDIS (Reporting Intellectual Capital to Augment Research, Development and Innovation in SMEs), Report to the Commission of the High

Level Expert Group. EUR 22095. Retrieved 11 August 2009, from http://www.madrimasd.org/proyectoseuropeos/documentos/doc/ricardis.pdf.

Figel, Ján (2006) 'EIT: a new model for the knowledge triangle'. Retrieved 26 July 2009, from http://ec.europa.eu/commission_barroso/figel/speeches/docs/06_04_25_EIT_stakeh_en.pdf

Forum on University-based Research (2005) *European Universities: Enhancing Europe's* EU High Level Group (2004) *Facing the challenge: The Lisbon strategy for growth and employment*. Chaired by Wim Kok. Retrieved 28 July 2009, from http://ec.europa.eu/growthandjobs/pdf/kok_report_en.pdf

Other Publications

Canibano, L., Sánchez, M.P., García-Ayuso, M. And Chaminade, C. (2002) *MERITUM PROJET. Guidelines for managing and reporting on intangibles (Intellectual Capital Report),* Vodafone Foundation: Madrid.

Committee on Facilitating Interdisciplinary Research (2004) *Facilitating Interdisciplinary Research*, National Academy of Sciences, National Academy of Engineering, Institute of Medicine, USA. Retrieved 31 July 2009, from http://books.nap.edu/openbook.php?record_id=11153andpage=2.

Garfield, E. (1979) *Citation Indexing. Its theory and application in science, technology and humanities*, New York: Wiley.

Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P. and Trow, M. (1994) *The New Production of Knowledge*, London: Sage.

Hazelkorn, E. (2008) **'**Learning to Live with League Tables and Ranking: The Experience of Institutional Leaders', *Higher Education Policy*, vol. 21, no. 2, pp. 195-215.

Hazelkorn, E. (2009) 'The Impact of Global Rankings on Higher Education Research and the Production of Knowledge', *UNESCO Forum on Higher Education, Research and Knowledge Occasional Paper N°16,* Paris.

Hirsch, J. E. (2005) 'An Index to quantify an Individual's Scientific Research Output', *Proceedings of the National Academy of Sciences (PNAS)*, vol. 102, no. 46, pp. 16569-16572. Retrieved 11 August 2009, from http://www.pnas.org/cgi/reprint/102/46/16569.pdf.

Jobbins, D. (2005) 'Moving to a Global Stage: A Media View', *Higher Education in Europe,* vol. 30, no. 2, pp. 137-145.

Locke, W. D., Verbik, L., Richardson, J. T. E. and King, R. (2008) *Counting what is measured or measuring what counts? League Tables and the impact on higher education institutions in England*, Circular 2008/14, Bristol: Higher Education Funding Council for England. Retrieved 11 August 2009, from http://www.hefce.ac.uk/pubs/hefce/2008/08_14/.

MacRoberts, M. H., and MacRoberts, B. R. (1996) 'Problems of citation analysis', *Scientometrics*, vol. 36, pp. 435–444.

Moed, H. F. (2009) 'New developments in the use of citation analysis in research evaluation', *Archivum Immunologiae et Therapiae Experimentalis (Warsawa),* vol. 57, pp. 3-18.

OECD (2002) *Frascati Manual*, Paris.

Price, D. J. D. (1970) 'Citation measures of hard science, soft science, technology, and nonscience', in Nelson, C.E. and Pollock, D.K. (ed.), *Communication among scientists and engineers*, Heath Lexington Books, Lexington, Mass., pp. 3–22.

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

Spaagen, J., Dijstelbloem, H. and Wamelink, F. (2007) *Evaluating Research in Context. A Method for Comprehensive Assessment*. 2nd edition. Consultative Committee of Sector Councils for Research and Development (COS): the Netherlands.

Stokes, D. (1997) *Pasteur's Quadrant: Basic Science and Technological Innovation*, Brookings Institution Press, Washington D.C.

Usher, A. and Savino, M. (2006) *A World of Difference: A Global Survey of University League Tables*, Canadian Education Report Series, Educational Policy Institute: Toronto. Retrieved 11 August 2009, from http://www.educationalpolicy.org/pdf/World-of-Difference-200602162.pdf.

Van Raan, A. F. J. (2005) 'Fatal Attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods', *Scientometrics*, vol. 62, no. 1, pp. 133-143.

Webography

Academic Ranking of World Universities, Shanghai Jiao Tong University, Retrieved 22 July 2009, from http://www.arwu.org/rank/2007/ranking2007.htm.

Classifying European Universities, Centre for Higher Education Policy Studies, University of Twente, Retrieved 22 July 2009, from www.utwente.nl/cheps/research/projects/ceihe/.

Design and Testing the Feasibility of a Multidimensional Global University Ranking. Retrieved 22 July 2009, from http://ec.europa.eu/education/programmes/calls/3608/index_en.html.

Elsevier *Scopus*, Retrieved 11 August 2009, from http://www.info.scopus.com/overview/what/.

ERIH (European Reference Index of the Humanities), European Science Foundation. Retrieved 11 August 2009, from http://www.esf.org/research-areas/humanities/research-infrastructures-including-erih/erih-initial-lists.html.

European University Association, *Institutional Evaluation Programme*, Retrieved 22 July 2009, from http://www.eua.be/events/institutional-evaluation-programme/home/.

Excellence in Research for Australia (ERA) journal lists for Physical, Chemical and Earth Sciences (PCE) and Humanities and Creative Arts (HCA). Retrieved 11 August 2009, from http://www.arc.gov.au/era/journal_list.htm.

International Observatory on Academic Rankings and Excellence, Retrieved 22 July, from http://www.ireg-observatory.org/index.php?option=com_frontpageandItemid=1.

Leiden World Ranking, Centre for Science and Technology Studies, University of Leiden, Retrieved 22 July 2009, from http://www.cwts.nl/ranking/LeidenRankingWebsite.html.

OECD Feasibility Study for the International Assessment of Higher Education Learning Outcomes (AHELO), Retrieved 22 July 2009, from http://www.oecd.org/document/22/0,3343,en_2649_35961291_40624662_1_1_1_1,00.html

Performance Ranking of Scientific Papers for World Universities, Higher Education Evaluation and Accreditation Council of Taiwan, Retrieved 22 July 2009, from http://ranking.heeact.edu.tw/en-us/2008/page/Background

QS World Universities (2008) Retrieved 12 August 2009, from http://www.topuniversities.com/worlduniversityrankings/%20methodology/simple%20overview/.

Ranking System Clearing House, Institute of Higher Education Policy, Retrieved 22 July 2009, from http://www.ihep.org/Research/rankingsystemsclearinghouse.cfm.

Ranking Web of World Universities, Cybermetrics Lab CSIC. Retrieved 22 July 2009, from http://www.webometrics.info/.

The Thomson Reuters Journal Selection Process, Retrieved 11 August 2009, from http://thomsonreuters.com/products_services/science/free/essays/journal_selection_process/.

Times QS World University Rankings. Retrieved 22 July 2009, from http://www.topuniversities.com/worlduniversityrankings/university_rankings_home/.

European Commission

EUR 24187 — Assessing Europe's University-Based Research - *Expert Group on Assessment of University-Based Research*

Luxembourg: Publications Office of the European Union

2010 — 148 pp. — 17,6 x 25,0 cm

ISBN 978-92-79-14225-3
doi 10.2777/80193

How to obtain EU publications

Publications for sale:

- via EU Bookshop (http://bookshop.europa.eu);
- from your bookseller by quoting the title, publisher and/or ISBN number;
- by contacting one of our sales agents directly. You can obtain their contact details on the Internet (http://bookshop.europa.eu)
  or by sending a fax to +352 2929-42758.

Free publications:

- via EU Bookshop (http://bookshop.europa.eu);
- at the European Commission's representations or delegations. You can obtain their contact details on the Internet (http://ec.europa.eu)
  or by sending a fax to +352 2929-42758.

In 2008, the European Commission, DG Research set up the Expert Group on Assessment of University-Based Research to identify the framework for a new and more coherent methodology to assess the research produced by European universities. There is no single, correct methodology. Any assessment of the quality of university-based research will have to take into consideration the multi-functional and diverse nature of universities, the diverse nature of disciplines, the level at which universities are assessed (i.e., identify the level at which knowledge is created and shared), the users' needs and assume an inclusive notion of research ranging from blue sky/curiosity-driven to user-led/practice-based research and interdisciplinary research. The key issue here is to design evaluations that are fit for purpose. While commercial rankings have positive and negative impacts onto the institutions, the students and other users, they tend to focus on only the world's top 100 out of the estimated 17,000 higher education institutions. Hence, there is a need to design flexible and multidimensional methodologies that will adapt to the diverse and complex nature of research, disciplines and of our universities.

K1-NA-24187-EN-N

Publications Office

ISBN 978-92-79-14225-3