Research universities and research assessment

May 2012

LEAGUE OF EUROPEAN RESEARCH UNIVERSITIES

Universiteit van Amsterdam • Universitat de Barcelona • University of Cambridge
University of Edinburgh • Albert-Ludwigs-Universität Freiburg • Université de Genève
Universität Heidelberg • Helsingin yliopisto (University of Helsinki) • Universiteit Leiden
KU Leuven • Imperial College London • University College London • Lunds universitet
Università degli Studi di Milano • Ludwig-Maximilians-Universität München • University of Oxford
Université Pierre et Marie Curie, Paris • Université Paris-Sud 11 • Université de Strasbourg
Universiteit Utrecht • Universität Zürich

If you can measure that of which you speak, and can express it by a number, you know something of your subject; but if you cannot measure it, your knowledge is meagre and unsatisfactory.

William Thomson, Lord Kelvin

Not everything that counts can be counted, and not everything that can be counted counts.

Albert Einstein

This paper has been written by Dr Mary Phillips, former Director of Research Planning at University College London and former member of LERU's Research Policy Committee, with the help of Dr Katrien Maes, Chief Policy Officer of LERU.

We thank the LERU Research Policy Committee, the LERU Vice-Rectors for Research and other individuals at member universities who provided valuable input for the paper and comments during the drafting process.

Research universities and research assessment

Executive summary

Research is at the core of research-intensive universities' multiple missions. Since university research is largely funded by public, often competitively allocated money in Europe, it is normal for it to be evaluated regularly as a matter of accountability.

In addition to external drivers, universities themselves are motivated by internal drivers for assessing their research effort. To remain vital and at the leading edge of a continually changing research landscape universities need ever more sophisticated research assessment tools suited to the task of strategically assessing research strengths and weaknesses for the institution as a whole or parts of it.

Externally and internally driven research assessment has incontestably become part and parcel of the research university enterprise. However, the explosion of various types of research assessment for a variety of users and purposes in recent years risks to create – or, one may argue, has already started to create – an obsession with measurement and monitoring, which may result in a "bean counting" culture detracting from the real quality of research and the boundless search for new knowledge.

The challenges and pitfalls for universities engaging in research assessment are numerous. The demands of producing large quantities of data require sophisticated HR and research management tools, databases and human expertise. Often different data are required for different types of assessment, resulting in unnecessary duplication of effort. Moreover, funding regimes tend to create incentives which may tempt universities or researchers to behave in certain ways, sometimes with unfortunate consequences.

Other challenges relate to the variety of users and purposes of research assessment. Exercises such as the British RAE/REF, the French *Initiatives d'Excellence* (Idex) and the German *Excellenzinitiative* are all linked to funding decisions. New national research evaluation agencies (e.g. AERES in France and ANVUR in Italy) have been set up to administer new assessment regimes. All have been the subject of intense debate and disagreement.

Some of this contestedness is due to the fact that research assessment can be performed with different methodologies. One of them, peer review, is largely regarded as a fair and effective method for assessing research potential and output even if its weaknesses are well-known. Bibliometrics provides a cost-effective alternative to peer review, but also has considerable drawbacks. As neither of them is perfect, bibliometrics may be used as a complement to peer review. Importantly, both need to be used with wisdom, discretion and the rigorous application of human judgement.

In recent years the European Commission has started to look at university research assessment and having now focused its attention on a related but different subject, that of university rankings, it is funding the development of a new ranking mechanism called U-Multirank. Although U-Multirank is intended as a user-driven, multi-dimensional (i.e. not just research) benchmarking tool, LERU has serious concerns about the lack of reliable, solid and valid data for the chosen indicators, about the comparability between countries, about the burden put upon universities to collect data, and about the lack of 'reality-checks' in the process.

A relatively new development in research assessment is the use of impact, which refers to results that are relevant beyond the research discipline and can be evidenced in benefits delivered to the economy, society, culture, public policy, etc. Impact is playing a considerable role in the new UK framework (REF) and in a new US government initiative called STARMETRICS, but the use of impact is also not without controversy.

Whatever the pros and cons, whatever the methodologies used, universities, governments and research funders alike need to "assess assessment", evaluating what works in different research environments, applying lessons learned rigorously but sensibly, and enabling informed decisions on the basis of valid and reliable evidence. Indeed, one of our main messages in this paper is that research assessment needs to be understood correctly and applied sensibly.

For universities, such understanding entails that assessment should reflect research reality and the needs of those involved. Goals, processes and criteria used should be defined clearly and transparently. Good assessment probably requires a suite of methodologies and most certainly good data management. Above all, universities should stand firm in defending the long-term value of their research activity, which is not easy to assess in a culture where return on investment is measured in very short time spans.

Governments, funders and others that seek to assess university research should recognise the broader role of universities in society and the long-term value they bring. They should work together to ensure that the information required from universities for assessment purposes is collected in a consistent manner which allows comparisons to be made between universities nationally, across Europe and if possible, across the world.

Introduction

1. Should university research be assessed? LERU's view is that university assessment or evaluation is an important and integral part of the university enterprise. Recognising all the arguments for and against, it wishes to contribute the views of research-intensive universities to achieve a better understanding of research assessment. In recent years we have witnessed an explosion of many types of internal and external research assessments for a variety of users and for a range of purposes, in which universities may choose or are obliged to engage. LERU feels therefore that a need has arisen to analyse what LERU universities consider to be the rationale for assessment, the internal and external drivers, the users and usages, the "pros and cons" and pitfalls, the variety of metrics used, different national and international models and emerging trends such as the increasingly important area of "impact". The paper also briefly considers international rankings of universities because, although assessment and rankings serve distinct purposes, some of the metrics used overlap. This is especially timely with regard to the development and implementation of U-Multirank by the European Commission. The paper concludes with recommendations for the users of research assessment, namely universities and researchers themselves, as well as governments, policy makers and funders of research.

2. In 1873 Benjamin Disraeli, the then British Prime Minister, told the House of Commons that a university should be a place "of light, of liberty and of learning". A 21st century understanding of this admonition might be that universities should engage in both education of students and enlightenment through research, both activities undertaken in an environment free from external constraints and interference. Yet most universities, especially those in Europe, are publicly funded and thus, in a very real sense, answerable to governments and the public, whose taxes to a large extent fund them. However, the notion that a university's activities should be assessed, appraised, benchmarked or ranked - an anathema to many academics and scholars - is relatively recent, entering the higher education (HE) lexicon in the last 30-40 years.

3. There is now an undoubted and growing recognition of the need for research intelligence and performance management frameworks and metrics to enable universities to assess and objectively benchmark their research activity. There is also widespread dissatisfaction with the tools currently available to integrate information from disparate systems, as well as an appetite for a more sophisticated approach. Despite a massive amount of work being undertaken globally in this regard over the past few decades, there is little consensus about the best approach (EC, 2010b). There is no simple answer. It is unlikely that one holistic form of assessment can be developed that can address all aspects of research in every academic environment.

4. Today, the area of research assessment is considered to be such a vital aspect of any university's activity, a number of commercial companies are developing sophisticated tools, mostly based on publications outputs (bibliometrics) to aid overstretched administrators and senior academics to conduct in-depth analyses of their faculty and benchmark the results against national and international competitors (e.g. Academic Analytics' Faculty Scholarly Productivity Index (FSPI), Elsevier's SciVal, Thomson Reuters' InCites). A number of universities have also developed tools (e.g. CWTS at Leiden University).

Challenges in assessment

5. Assessment or benchmarking of research usually involves the compilation of data relating to
   a) inputs, such as competitively won research income, human and physical infrastructure, and research environment,
   b) outputs, including publications, citations, PhDs produced, commercialisation (e.g. patents, spin-outs, licences, venture-capital),
   c) outcomes, i.e. longer-term societal and economic impacts.

   It could be argued that the latter two, outputs and outcomes, are more valid measures since they indicate what the research has actually achieved. In some cases it may also involve data on the "process" of research - how research is conducted (governance); the effectiveness or efficiency of the HR or research offices; technology transfer capability etc.

6. Any assessment system must be sensitive to the possibility of generating perverse incentives that promote unhelpful and dubious management practices, more at home in the realms of football administration than academe. It needs to be mindful of disciplinary differences and have a long enough time frame - at least five years - to be meaningful. There are inherent dangers which need to be weighed against the benefits.

7. The demands of producing large quantities of data may create or extend a culture of regulations, instructions, lists of good practice etc., marked by an obsession with measurement and monitoring, all of which might well detract from creative freedom, flexibility and productivity.

8. At face value, all of the indicators used should be readily accessible. There are however a number of significant issues which make the collection of data and meaningful comparisons between institutions, even in the same country, challenging. These include the lack of clear and shared definitions of a number of metrics, even such seemingly elementary ones as what is meant by a "researcher".

9. There also needs to be an understanding that merely "bean counting" is not enough. For example, just counting the number of PhDs produced is not a true reflection of output. There needs to be an assessment of quality. In the UK, for example, many universities have recognised that in most science disciplines, four-year PhDs (often with a taught element) provide better quality research training than the traditional three-year model. With a limited budget, this will result in fewer (but arguably better) PhDs trained. Such quality related issues are not reflected in most assessment regimes.

10. A further example in relation to PhD training is provided by a European evaluation agency which, wishing to establish an index to quantify "attractiveness" of research laboratories, took as a criterion the number of foreign PhD students and post-docs, but only from industrialised countries. Such students, coming from countries in which the facilities to conduct their research project were probably available, chose foreign laboratories on the basis of high scientific standing and thus attractiveness. In that sense, this defined index of attractiveness is relevant. However such a strategy also has a perverse effect. It ignores other goals, such as helping developing countries improve and grow their science base, and the positive influence this may have in determining the direction of science in those countries.

11. In many instances institutions have allowed the demands of external stakeholders to determine the data and the data definitions they collect and measure. With no overarching and consistent approach, it is not unusual for institutions to submit different information for the same data point to various external data gathering exercises. Moreover, benchmarking requires not only an institution's own data but also proprietary data (e.g. held by funders) and data held by third parties, all of which can be inconsistent or difficult to access. There often is a duplication of effort, frequently involving manually intensive systems.

The university perspective - internal drivers

12. In addition to governments (and tax-payers) wanting to be confident that the significant investments made in university research are being used effectively, there are a number of other reasons why universities themselves might want to undertake some form of assessment (or benchmarking). Although focused on research, such assessments inform the interlinked missions of universities, i.e. education, research, innovation and societal impact. Assessment, here defined as the process of gathering, collating, analysing and synthesising specific information to provide informa-

tion and intelligence as part of an evaluation, is not to be confused with ranking (see paragraphs 24-28). The rationale for assessment within a university might be to:

- rigorously gauge research output, quality and impact, thus ensuring future allocation of funds, improving performance and maximising return on investment;
- provide the academic community with an opportunity to receive topical and versatile international peer feedback enabling identification of strengths and areas to be developed;
- promote the recognition of the university's research potential;
- inform strategic planning in specific subject areas, thereby facilitating investment in accord with research strengths or developing important new areas, or to expose weaknesses in need of remediation (or possible disinvestment);
- identify and track individual accomplishments;
- recruit, retain and reward top performers;
- track (and possibly reward) departmental/faculty performance and leadership;
- find and foster productive collaborations, including international ones, especially with rapidly developing economies;
- benchmark against genuine institutional peers (and potential peers) to assist positioning in increasingly internationally competitive academic environments.

The rationale for all of the above would be supported and welcomed by most senior university academics and administrators.

The users of assessment and benchmarking

13. In addition to universities and governments, other groups of stakeholders have an interest in gauging the quality and productivity of universities and of individual researchers/groups within those universities. These include charitable foundations wishing to support the best research, either through competitive funding programmes or philanthropic donations; commercial sponsors looking for appropriate academics to undertake commissioned research or venture capitalists seeking potential investment opportunities through university spin-outs; researchers looking to relocate to environments where their research potential might be more effectively channelled; students, both undergraduate or postgraduate, national and international, looking for environments where they will be exposed to or taught by outstanding researchers in their areas of interest; external researchers/groups seeking collaborators, increasingly across national boundaries. Last but not least, individual researchers and groups, those working at the "coal-face", who may not have a broad knowledge or understanding of what is happening within their university as a whole, also need to have the information fed back to them to gauge how they perform relative to their peers. They can use this information to set aims and find future partners. The evaluation can thus also be conducted as an enhancement-led process, rather than as a mere benchmarking activity, providing useful information to enrich and strengthen the research of academic communities.

14. Whilst the needs of these diverse "stakeholders" may overlap, some have specific requirements which must be addressed. However, whatever the nature of the information and data collected, it must be timely, robust, accurate, transparent and verifiable.

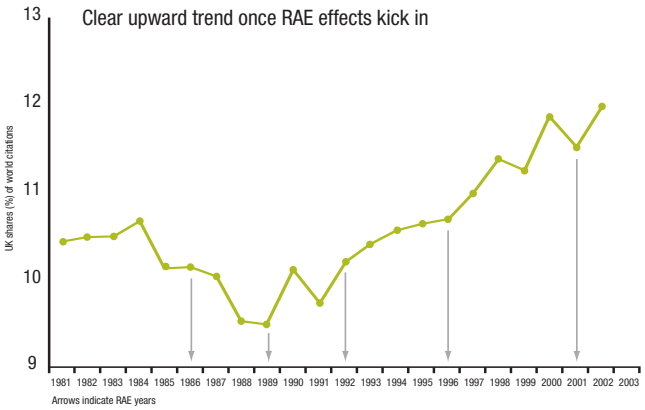
Assessment models in different national settings

15. In the UK the first official exercise to assess university research took place in 1986 under the Thatcher Government. It was conducted by the University Grants Committee, a predecessor of the present Higher Education Funding Councils. The purpose of the exercise was to determine the allocation of funding to UK universities at a time of tight budgetary restrictions, the latter point being a crucial driver.

16. This process, known as the Research Assessment Exercise (RAE), has evolved and changed over two and a half decades into the current Research Excellence Framework (REF), which will come to fruition in 2014. These assessment exercises are used by the Funding Councils to allocate QR (Quality Related) block grants selectively, according to a formula which rewards excellence, in the various units of

assessment. Although funding supports excellence wherever it is found, the process has resulted in the focusing of funds in the research-intensive universities. Despite the fact that these exercises have become increasingly costly both to the Funding Councils and to universities, there is good evidence that the block grant and the assessment process which allocates it, have played a key role in achieving, since the 1980s, the resurgence of UK research of the highest quality (Evidence Ltd, 2000; HEFCE, 2000; HEPI, 2004). Successive RAE cycles have driven improvements not only in the performance of research-intensive institutions but also "at all grades and across subject areas" (Evidence Ltd, 2005). The block grant therefore represents a highly successful allocation of research funding. The key attribute of QR funding is that is not hypothecated, i.e. its flexibility enables research leaders and senior administrators to make strategic decisions about its use and distribution, facilitating the emergence of new priorities and areas of enquiry. Data also show that there is a correlation between QR and other funding streams including industrial ones. The graph below illustrates the increase in the UK's share of world citations as the RAE takes effect over the years[1].

However the RAE has not been without criticism. In 2004 Lord May of Oxford, at the time President of the Royal Society, described the "pathologies of the RAE" (May, 2004). Whilst not detracting from the need for external review of universities, he pointed out the difficulty of assessing research which often cuts across the rigid boundaries of the units of assessment, both within a university or involving external collaboration. He recommended a longer period between assessments and more flexible approach which recognised "what actually goes on in the creative process". Although this model appears to have had a positive effect in the UK, it may not be appropriate for other countries, especially those in Europe, where the university and funding systems are somewhat different.



Research assessment has led to an increase in UK share of world citations

17. Assessment processes have been introduced in other European countries and around the world, all having depended to some extent on trying to objectively quantify scholarly activities (EC, 2010b). In Australia, for example, the Excellence in Research for Australia (ERA) was launched in 2008. This replaced the somewhat controversial RFQ or Research Quality Framework, a process very similar to the UK's RAE which was aborted on the grounds that its design was cumbersome, it lacked transparency and incurred high costs of implementation. The ERA will be more streamlined, with a greater use of indicators or metrics and a lesser burden on researchers, institutions and expert assessors. Any attempt to measure broader "impact" has been dropped. Following a complete cycle through all the panels over three to four years, the government is likely to attach funding to outcomes which will determine some or all of the university block grants for infrastructure, training and research.

18. Assessment has come to the fore in Germany through the Excellence Initiative, which, introduced in 2005, aims at promoting top-level research at universities and raising their international profile. The Excellence Initiative is implemented jointly by the German Research Foundation (DFG) and the German Council of Science and Humanities (*Wissenschaftsrat*) and comprises three funding lines: Graduate Schools to promote young academics, Clusters of Excellence to promote top-level research, Institutional Strategies in the form of projects in support of top-level research at universities[2]. The funding (€1.9 billion) from the federal and state governments was awarded to nine universities in 2006-2007 on a competitive basis with the decisions being made by a number of expert committees. There is some ambivalence amongst stakehold-

1 Source: Higher Education Funding Council for England, www.hefce.ac.uk

2 See http://www.wissenschaftsrat.de/1/fields-of-activity/excellence-initiative/

ers as to the benefits of the initiative. Proponents highlight benefits such as the differentiation and strategic focusing of universities and the cooperation with non-university research institutions and business partners (Kleiner, 2010). Critics emphasize among other points that the funding model generates competitive disadvantages for non-experimental fields of research and that funding is directed towards research at the expense of teaching. As a result, some of the criticisms have been taken into consideration for the second round of the competition starting in 2012. The amount of funding allocated for research clusters now spans a greater range, making this funding line more attractive for the arts and humanities. Furthermore the Federal Ministry of Education and Research has initiated an independent programme to foster quality in teaching.

19. Assessment of relevant quality performance indicators, established during the creation of the Excellence Initiative, is increasingly taking centre stage in the German higher education system. A recent study published by Forschungszentrum Jülich (Mittermaier, 2011) shows the number of publications and their development as significant indicators for research output and visibility of academic institutions, having found a strong correlation between the nine universities funded during the first period and an above average increase of publications as compared to other German HE institutions.

20. In France, *the Initiatives d'Excellence* (Idex) is part of a major shake-up of French higher education[3]. The plan, controversial since it runs counter to the egalitarian tradition in French higher education, is designed to establish a number of world class universities (currently eight) capable of competing internationally for the best students and academics. It requires non-selective universities, highly selective *grandes écoles* and independent research institutions such as CNRS, to work together in larger units, in exchange for massive new investment. Funding is competitive and outcomes are determined by an international jury, which bases its decisions on four criteria, i.e. research excellence, training and innovation capacity, national and international partnerships, and strategic management capability. However, there is some debate amongst academics as to whether bigger will necessarily be better. Clearly, gathering large numbers of researchers into a single institutional entity will increase the number of publications, citations and other measures of output, thus enhancing the visibility of the new "unit", however only time will tell as to whether there is a real increase in quality and productivity.

21. In addition, a new assessment agency was set up in France in 2006. AERES (*Agence d'Evaluation de la Recherche et de l'Enseignement Supérieur*) is designed as an independent administrative authority, aimed at leveraging research and higher education in France in agreement with the European recommendations and the Bologna process[4]. The agency's scope is broad, covering the evaluation of all types of research organisations, universities and other higher education and research institutes, as well as ANR (National Research Agency). It addresses both the assessment of research activities as well as education programmes and curricula. It has independent administrative authority status, allowing it to carry out its missions without having to bow to any pressure or subordination from governmental authorities, evaluated parties or any other stakeholders. AERES makes all of its evaluation reports accessible to the public, publishing its methods and procedures in complete transparency and calling on foreign experts with diverse backgrounds who are chosen depending on what is being evaluated.

22. In Italy, in the last three years an increasing part of the state support to universities (5% initially and now 7%) is distributed based on research parameters. An ad hoc agency (*Agenzia Nazionale per la Valutazione dell'Università e della Ricerca* - ANVUR) has been established and a new evaluation exercise is now under way. Evaluation panels include a significant representation of scientists working abroad. These steps represent a fairly dramatic change relative to previous practices in university funding.

23. In recent years the European Commission has started looking at university-based research assessment in Europe and at the growing obsession with a related but different issue, namely that of ranking universities. These topics are dealt with in the next section.

3 See http://investissement-avenir.gouvernement.fr/content/action-projets/les-programmes/centres-dexcellence (in French), also http://www.insidehighered.com/news/2011/10/07/france_attempts_to_create_league_of_elite_universities, and http://news.sciencemag.org/scienceinsider/2012/02/france-picks-five-more-university.html

4 See http://www.aeres-evaluation.fr/Agence/Presentation/Profil-de-l-agence

The EU perspective - from assessment to rankings

24. In 2002 the European Union adopted the Lisbon Agenda (Strategy) as an economic action and development plan. Its aim was to make the EU "the most competitive and dynamic knowledge-based economy in the world" by 2010. It followed that there was a need in Europe for a research base that was cutting-edge and second-to-none. This would be driven by Europe's universities and research institutes, where a large proportion of research, especially "blue-skies" research, is undertaken. This Agenda, which had made at best very modest progress by 2010, is now carried forward in the Europe 2020 Strategy and its flagship initiatives among which the Innovation Union (EC, 2010a). LERU (2010a, 2011) has repeatedly described the role of research-intensive universities in contributing to economic progress and innovation and has suggested ways in which the EU can help to build and support world leading research universities in Europe.

25. In such a context, in which economic performance is linked to research performance, the European Commission convened an expert group on assessment of university-based research in 2008. Its mandate was to consider existing assessment systems and to identify indices that should/could be used in the future. The expert group discovered that "users" were manifold and diverse, from HE senior management, governments and government funders, to other academic research organisations, peer HE institutions, individual researchers and students, employers, sponsors and private investors, media and the public. Looking at the strengths and weaknesses of existing methodologies, the expert group concluded that 1) there was no single perfect indicator and most objective indicators were proxies; 2) indicators must be fit for purpose and verifiable; 3) the variety of ways in which different disciplines publish and disseminate can positively or negatively affect the choice of indicator, especially using bibliometric data.

26. Not unexpectedly, in their final report (EC, 2010b) the expert group suggested that any assessment should combine quantitative, indicator-based information with qualitative, peer- and end-user review. They also recommended more consideration of the "knowledge cluster", defined as a group of researchers sharing a common field of investigation. In many universities these would be departmentally based (e.g. history, biology, etc.), but in some areas they may be aggregated at an intra- or inter-institutional level. In addition, the report stressed the need to a) improve bibliometrics to cover all disciplines; b) devise methods to capture "impact", interdisciplinary and collaborative research; and c) improve HE institutions' capacity to capture or collect, maintain, analyse and disseminate standardised data so as to make comparisons nationally and internationally.

27. The expert group also analysed global university rankings, noting that they were becoming a popular way of assessing university-based research[5]. They warned against rankings or assessment systems seeking to "compare whole universities on the basis of an aggregated score and which lack validation through expert peer assessment" (EC, 2010b, p.16). They concluded that, in the absence of 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" (ibidem).

28. Be that as it may, in the face of mounting political pressure to produce a Eurpean ranking system, the Commission set up a feasibility study to develop a global but "made-in-Europe" ranking system of universities. "U-Multirank" was to be a new, performance-based, multi-dimensional tool to capture, in quantitative terms, the nature and performance of individual universities in Europe and beyond[6]. Aiming to radically improve the transparency of the sector, U-Multirank takes into consideration five dimensions of universities' activities, namely teaching and learning, research, knowledge exchange, internationalisation and regional engagement. Having been received pos-

5 University rankings, both national and international, are related to, but somewhat different from, assessment or benchmarking. Rankings or "league tables" are lists of institutions of higher education, ordered by combinations of factors usually relating to both teaching and research (and increasingly "impact"). A global phenomenon, they are conducted by universities or institutes (SHJT-ARWU, CWTS-Leiden), by commercial entities (Thomson Reuters-THE, QS, Webometrics) and by governments (U-Multirank).

6 More information on U-Multirank, including the final report of the feasibility study, is available at http://www.u-multirank.eu/. U-Multirank is not to be confused with U-Map. The latter is a classification tool to describe different types of universities and what they do, whereas U-Multirank aims at ranking, i.e. showing how well universities do certain things. See http://www.u-map.eu/

itively by the EC, the project has gone into a next phase of development and is expected to publish its first attempts at implementation in 2013. Not unsurprisingly, the initiative is taken up in the recent EC Communication on the modernisation of Europe's HE systems (EC, 2011), which covers issues such as governance, autonomy and the improvement of higher education.

LERU and rankings

29. All systems of university rankings have been subject to varying degrees of criticism (EUA, 2011; LERU, 2010b). These include problems with both peer review and bibliometrics (see next sections) as well as the blanket application to the whole university when there may be wide variations across the institution. In addition, the performance indicators used are mostly proxies (i.e. indirect and often inappropriate measurements) and are weighted depending on the importance the compiler ascribes to the metric. The capacity of rankings to measure the true value of universities to society remains to this day poor and yet rankings remain powerful drivers of often undesirable behaviour by universities, policy makers and governments. Therefore they risk doing more harm than good (LERU, 2010b).

30. As for U-Multirank, LERU was involved in the feasibility study, but serious concerns about the project have lead LERU as an organisation to decide not to engage further. Our main concerns relate to the lack of good or relevant data in several dimensions, the problems of comparability between countries in areas such as funding, the fact that U-Multirank will not attempt to evaluate the data collected, i.e. there will be no "reality-checks", and last but by no means least, the enormous burden put upon universities in collecting the data, resulting in a lack of involvement from a good mix of different types of universities from all over the world, which renders the resulting analyses and comparisons suspect.

31. Whereas LERU is uncomfortable with the way in which rankings have evolved for the reasons given above, we are very much convinced of the need to evaluate universities' research, both by universities themselves who need to evaluate their researchers, research programmes and research strategies, and by others who have a responsibility to evaluate whether public money invested in research is well spent. In the next section we return to assessment, looking at two commonly used tools for assessment.

Assessment methodology

32. Many of the processes used to assess research are based on two methodologies, peer review and bibliometrics. The following sections look at the advantages and disadvantages of both.

Peer review - the pros and cons

33. Peer review is defined as a process of self-regulation by a profession or a practice of evaluation involving qualified individuals within the relevant field. Peer review methods are employed to maintain standards, improve performance and provide credibility. It is probably still the most recognised way of assessing research for both the distribution of funding and judgements about publication of results. However it is not without its detractors. Dr Richard Horton, Editor of the Lancet, wrote in 2000 that "we portray peer review to the public as a quasi-sacred process that helps to make science our most objective truth teller. But we know that the system of peer review is biased, unjust, unaccountable, incomplete, easily fixed, often insulting, usually ignorant, occasionally foolish, and frequently wrong" (Horton, 2000). This conclusion is clearly extreme but it does contain some element of truth.

34. The selection of peers may introduce bias into the system, sometimes deliberate but mostly inadvertent, and their judgements are subjective and can be inconsistent, often swayed by group bias. The process tends towards conservatism and stifles innovation, accepting the status quo. Thus new ideas which challenge conventional wisdom do not always fare well unless reviewers are very enlightened. It probably disadvantages interdisciplinary research and is increasingly burdensome for the reviewers. It is also very costly. Many studies have sought to investigate whether peer review discriminates against particular groups such as young researchers, women, those from less prestigious institutions, non-native English speakers and those with unconventional views from outside the mainstream. Taken as a whole, the results of such studies are inconclusive (RIN, 2010).

35. Whatever its failings and weaknesses most researchers believe that review by their peers is prob-

ably the fairest and most efficient method of assessing past work and future potential. There have been numerous attempts by funding bodies and publishers to improve the system such as taking full advantage of advances in digital technologies, increasing the pool of potential reviewers especially seeking opinions from outside the country to avoid parochialism and cronyism, and using standardised formats for reviews. However, its essence is still one of informed opinion. In some disciplines, where bibliometric methods are less well established, the arts and humanities for example, peer review remains the most effective form of assessment.

36. A comprehensive European Peer Review Guide was recently published by the European Science Foundation. It was anticipated that this very useful Guide "should serve to benchmark national peer review processes and to support their harmonisation, as well as to promote international peer review and sharing of resources" (ESF, 2011).

Can we trust bibliometrics?

37. Bibliometrics is usually credited as originating with the work of Garfield et al. in the 1950s and the subsequent development of citation indexing and search tools such as ISI and SCI. However, it has in fact been around since the 1900s with the work of James McKeen Cattell (an American psychologist and editor of Science in the period 1895-1944), who was looking at productivity in his own discipline. It is now defined by the Oxford English Dictionary as "the branch of library science concerned with the application of mathematical and statistical analysis to bibliography; the statistical analysis of books, articles, or other publications. Alternatively it is "the discipline of measuring the performance of a researcher, a collection of articles, a journal, a research discipline or an institution". Thus bibliometrics can be employed to assess and rank the research outputs of individuals, institutions and countries. It is used in evaluation and decision making by universities and their senior academic administrators, policy makers and researchers themselves. A powerful case is made for bibliometrics by Van Raan (2005), arguing that "advanced bibliometric methodology provides the opportunities to carry out effective evaluations with low burdens for the objects of the evaluation".

38. Bibliometric outputs/outlets differ between disciplines. Those in the natural and life sciences publish mostly journal articles; engineers publish journal articles and conference proceedings; social scientists and humanities scholars focus on journal articles, book chapters, monographs and books. These differences need to be taken into account in assessments in these areas. Of course those working in the arts (art, music, dance, drama) and architecture may not "publish" at all, but rather produce artefacts and the like, the research elements of which are challenging to assess.

39. Bibliometric indicators vary between disciplines. There are hierarchies of outlets, with medicine for example tending to give weight to impact factors for journals, while this is less important for the social sciences. The scope of research will affect the type of journal published in: although fundamental science is universal and therefore more appropriate in international journals, some research is more important locally or regionally and it therefore makes sense to publish in journals that are more likely to be read at a national level. The language of science is mostly English but in some disciplines publication is more appropriate in the national language. The time span is also a factor: in some areas of science the pace of change is very fast and thus papers from three to five years before are less likely to be cited.

40. There are a number of well documented pitfalls about the blanket application of bibliometrics such as how to cope with papers having multiple authors, particularly in certain fields (e.g. genomics or high-energy physics); variations in how the names of researchers and their institutions appear; over-citation of reviews and articles describing methodologies; self-citation; the different patterns of publication and citation in different disciplines. Additional issues arise from the fact that in some areas, in particular biomedicine, significant publications appear in the "grey" literature - i.e. reports from governments, NGO's, etc. - and increasingly in electronic and open-access journals. Most providers of bibliometric data take these into consideration. Despite the costs of peer review compared to bibliometrics, it is generally agreed that bibliometrics should supplement and complement, not replace, the former, providing a quantitative basis for decision making, but be used with discretion and the rigorous application of human judgement.

Impact - a new dimension to assessment

41. Increasingly funders and other stakeholders, whether public or private, require data, to justify future investment. They demand evidence that the resources provided to academic researchers are having a broader impact (changes in public policy, improvements in treatments, environmental impacts, public engagement, etc.) beyond the purely intellectual contribution to advancing the discipline. Quantitative metrics are being developed but the area is very challenging especially in some disciplines. It might be that in many cases impact can only be described qualitatively. However, even in the most esoteric of areas, where application to the problems of the modern world is not obvious, such research can help us understand where we came from and what has influenced our development.

42. It should also be remembered that one of the most important impacts of research is to nourish, refresh, update and rethink the other major aspects of university life, in particular the education of the next generation.

Impact - will the UK model work?

43. The UK's REF (Research Excellence Framework) now includes a significant score (20%) for impact, defined as benefits to the economy, society, culture, public policy and services, health, the environment, international development and quality of life. Such "impact" will be assessed against equally demanding standards as those of more traditional output measures. The two key criteria will be the "reach" of the impact - how widely has the impact been felt, and "significance"- how much difference was made to the beneficiaries. Universities will be asked to provide "case studies" in the various units of assessment, with a time frame for the occurrence of the impact of the order of five years, based on underpinning research conducted over the last twenty years. Although this approach is very worthy and appropriate, it is likely to be highly resource intensive and very new for many academics. In addition much research is collaborative, making it difficult to track and evaluate the relative contribution of individual academics or groups. It also represents a new challenge for the assessors of "impact" and panels will require specialist training to adequately judge the evidence presented.

44. Clearly identifying and quantifying impact will vary widely between different disciplines. In medical sciences, for example, the discovery and development of new drugs, devices or other therapeutic interventions will be readily quantifiable in terms of both human well-being and economic benefit, although the basic research on which the intervention is based may have taken place over decades. An often quoted example of the latter is the identification of monoclonal antibodies in the MRC's Laboratory of Molecular Biology by Milstein and Kohler in 1970 - work for which they were awarded the Nobel Prize in 1994. It took a further 30 years or more to see the development of drugs based on the technology, which now account for about one-third of the biotechnology healthcare market with applications in oncology, inflammation, organ transplant and more[7].

45. In 2010 the Higher Education Funding Council for England (HEFCE) carried out a number of pilot exercises to gauge the ease with which universities would be able to undertake reporting of impact for different disciplines and sizes of units. A report of their findings concluded that "in the main, the feedback confirmed that individual Pilot Institutions have derived much insight and learning from the exercise, which might be of significance to the wider community" (Technopolis, 2010). It also identified a number of shortcomings in the HEFCE guidance and provided recommendations for improvements in the REF proper. It also recognised the considerable costs likely to be incurred by institutions in preparing impact material.

The US approach to impact

46. A recent development in the USA has been the development of STAR METRICS[8]. This new initiative promises to monitor the impact of federal science investments on employment, knowledge generation and health outcomes. It is a multi-agency venture led by the National

7 For more discussion and examples of the long and often circuitous route that research follows to result in new goods or services see the LERU statement "Getting to grips with the competitive challenge" (2011) and the accompanying case studies available at www.leru.org.

8 STAR METRICS stands for Science and Technology for America's Reinvestment: Measuring the Effect of Research on Innovation, Competitiveness and Science. https://www.starmetrics.nih.gov/

Institutes of Health (NIH), the National Science Foundation (NSF), and the White House Office of Science and Technology Policy (OSTP). STAR METRICS will help the federal government document the value of its investments in research and development to a degree not previously possible. Together, NSF and NIH have committed $1 million for the programme's first year. Francis Collins, Director of NIH, maintains that "STAR METRICS will yield a rigorous, transparent review of how our science investments are performing. In the short term, we'll know the impact on jobs. In the long term, we'll be able to measure patents, publications, citations, and business start-ups"[9].

47. Data for the programme will come from research institutions that volunteer to participate and the federal agencies that fund them. Information will be gathered from the universities in a highly automated way, with minimal or no burden for the scientists and the university administration. STAR METRICS is based on a successful pilot programme that includes seven research institutions and is now being extended to more universities, with 85 already having expressed interest in taking part, representing 50% of NIH/NSF funding.

48. There are two phases to the programme. The first phase will use university administrative records to calculate the employment impact of federal science spending through the agencies' existing budgets and the American Recovery and Reinvestment Act. The second phase will measure the impact of science investment in four key areas: economic growth will be measured through indicators such as patents and business start-ups; workforce outcomes will be measured by student mobility into the workforce and employment markers; scientific knowledge will be measured through publications and citations; and social outcomes will be measured by long-term health and environmental impact of funding.

49. The initiative will in theory enable the federal government to justify spending on research to the US taxpayer, and funding agencies to locate experts and analyse gaps. Moreover, senior university administrators should be able to identify strengths and weaknesses, and researchers to locate others in their field and to ascertain how the latter are being funded. Since the programme is not mandatory, its success or failure will, as with U-Multirank, depend to a large extent on the collaboration and involvement of the majority of the research universities and the quality of the HR and finance records of those universities, since the process of data collection is entirely automated. Already anecdotal evidence suggests that a number of anomalies appear to be occurring. There is concern about coverage especially in disciplines that focus on highly selective and tightly focused conference proceedings, traditional journals being deemed to slow. In addition, it is thought that there may be perverse effects on young new investigators.

Impact and Europe

50. The Innovation Union is one of the flagship initiatives of the Europe 2020 Strategy (see paragraph 24). It contains 30 action points which aim to turn Europe into a world-class research and innovation performer. The dimensions of assessment likely to be considered in the allocation of research funding are excellence, impact and implementation. This approach is in marked contrast to that taken by the European Research Council. The ERC is universally acknowledged to have been hugely successful in supporting and promoting "frontier" research across the academic disciplines with "excellence" being the only criterion for funding. In her keynote address at the ERC's fifth anniversary conference, ERC President Helga Nowotny reaffirmed this approach and rebutted any suggestion that ERC funding should be more informed by impact (Nowotny, 2012). Interestingly, it should be noted that the ERC last year introduced "proof of concept" grants which aim to help existing grantees to take the first steps in bringing good ideas to market. Although the funding for this scheme is limited, it demonstrates that the ERC is not antagonistic to the idea of research being exploited, but rather that it should not be the driving force.

9 Source: www.nsf.gov/news/news_summ.jsp?cntn_id=117042

Conclusions

51. Whatever the pros and cons, and whatever processes are used, in a world where the costs associated with research are soaring, and there is a global market for talent, research assessment is here to stay. The task for governments and universities is to "assess assessment", i.e. to look at what works in different environments and research cultures globally, and to build on best practice where there is quantifiable evidence that the process leads to demonstrable improvements in productivity and impact. Within Europe, LERU will continue to inform, support and where appropriate, lead this debate.

52. Our main message is that research assessment needs to be understood correctly - what it says or does not say about research universities as institutions whose main purpose is to create new knowledge and deliver that knowledge to society through its teaching, research and societal missions (cf. also LERU's mission statement[10]). Research assessment therefore needs to be applied rigorously but sensibly, so that its varied users can make informed and justifiable decisions on the basis of valid and reliable evidence.

53. Most of all, universities should stand firm in defending the long-term value of their research activity to society, finding the right balance between the need to account for the use of public (and private) money invested in research and the pressure to measure and quantify virtually every aspect of the research enterprise, including some elements which are probably immeasurable. Case studies and public and media engagement are also valuable, while not necessarily quantifiable, ways to demonstrate societal value.

54. We end this paper with specific recommendations for the users of university research assessment, namely universities and researchers themselves, as well as governments, funders and other external agencies.

10 LERU's mission is to advocate education through an awareness of the frontiers of human understanding; the creation of new knowledge through basic research, which is the ultimate source of innovation in society; and the promotion of research across a broad front in partnership with industry and society at large.

LERU's recommendations

Recommendations for universities and researchers

55. From a university perspective, evaluation should reflect research reality and the needs and aspirations of those involved. Thus, senior administrators and academics must take account of the views of those "at the coal-face" of research when developing assessment criteria and indicators (as should governments, funders and other external agencies).

56. The assessment process should be as transparent as possible and the objectives explicitly defined. A balanced and comprehensive research assessment needs to include a suite of methodologies, appropriately reflecting inputs, outputs and longer-term impact, with a clear understanding of the limitations of each metric, especially at a discipline-specific level.

57. Universities can enhance the efficiency of research assessment by ensuring that the data on which it is based is as accurate, current and readily accessible as possible. Many universities struggle to maintain up-to-date HR and accounting systems and indeed to define exactly what is meant by a "researcher"[11] or to know exactly who has published what and where.

58. Researchers should be encouraged (or compelled) when publishing, to use a unique personal and institutional designation, and to deposit all publications into the university's publications database.

59. Similarly, information relating to grants awarded, PhDs trained, measures of esteem, commercial activities (patents, licenses, spin-outs etc.) and other input and output measures should be accurately assigned and preferably held in a central comprehensive research database. To maintain the trust and co-operation of the research community, senior administrators should also ensure that the information collected can be used for multiple purposes, both internal and external, to avoid duplication of effort.

Recommendations for governments, funders and other external agencies

60. Governments, funders and other external agencies should work together to ensure that the information required from universities for assessment purposes is collected in a consistent manner which allows reliable comparisons to be made between universities nationally and within Europe (and ideally internationally beyond Europe) and which does not overburden the institutions themselves. Again there needs to be a clear understanding of the objectives of such evaluations and transparency in the indicators used.

61. Any system which relies on an automated process, such as STAR METRICS, requires regular "reality checks", to ensure that the results are realistic and believable.

62. Governments and other external agencies should recognise that behind each index or evaluation criterion they include, there may a hidden objective/agenda that universities then implicitly espouse by encouraging their teams to reach the highest scores.

63. All those who seek to assess university research should recognise the broader role that these institutions have in society, and value the long-term benefits that universities bring. This is not to dismiss measurement of "value" but to undertake any quantification sensitively and validly taking these factors into account.

11 See point 4 and LERU (2010c) report *Harvesting talent: strengthening research careers in Europe*.

References

- European Commission. (2011). *Supporting growth and jobs - an agenda for the modernisation of Europe's higher education systems*. COM (2011) 567.
- European Commission. (2010a). *Europe 2020 flagship initiative Innovation Union*. COM (2010) 546.
- European Commission. (2010b). *Assessing Europe's university-based research.* Report by the expert group on assessment of university-based research.
- European Science Foundation. (2011). *European peer review guide - integrating policies and practices into coherent procedures.*
- European Universities Association. (2011). *Global university rankings and their impact.*
- Evidence Ltd. (2000). *The role of selectivity and the characteristics of excellence*.
- Evidence Ltd. (2005). *Impact of selective funding of research in England, and the specific outcomes of HEFCE research funding*. Report to HEFCE and the Department for Education and Skills.
- German Council of Science and Humanities [*Wissenschaftsrat*]. (2010). *Assessment criteria for institutional strategies - renewal proposals.*
- Higher Education Funding Council for England. (2000). *Fundamental review of research policy and funding: subgroup to consider the nature and purpose of HEFCE funding.* Final report.
- Higher Education Policy Institute. (2004). *What future for dual support?* Jiao Tong University.
- Horton, Richard. (2000). *Genetically modified food: consternation, confusion, and crack-up.* MJA 172(4) 148-149.
- Kleiner, Matthias. (2010). *Keine Etiketten auf Dauer.* In:duz - Deutsche Universitätszeitung- Magazin für Forscher und Wissenschaftsmanager. Heft 9/2010.
- League of European Research Universities. (2011). *Getting to grips with the competitive challenge.* Statement to the EU Heads of State and Government, January 2011.
- League of European Research Universities. (2010a). *Universities, research and the Innovation Union*. Advice paper nr. 5, October 2010.
- League of European Research Universities. (2010b). *University rankings: Diversity, excellence and the European initiative.* Advice paper nr. 3, June 2010.
- League of European Research Universities. (2010c). *Harvesting talent: strengthening research careers in Europe.* Position paper, January 2010.
- May, Robert. (2004). *Pathologies of the RAE.* The Journal of the Foundation for Science and Technology. June 2004.
- Mittermaier, Bernhard. (2011). *Publizieren Spitzen-Unis mehr?* In: duz - Deutsche Universitätszeitung- Magazin für Forscher und Wissenschaftsmanager. Heft 9/2011.
- Nowotny, Helga. (2012). Speech at the European Research Council fifth anniversary event, Brussels.

- Pasternack, Peer. (2008). *Die Exzellenzinitiative als politisches Programm - Fortsetzung der normalen Forschungsförderung oder Paradigmenwechsel?* In: Bloch et al. (Hrsg.). *Making Excellence. Grundlagen, Praxis und Konsequenzen der Exzellenzinititative*. Bertelsmann. 13-36.
- Research Information Network. (2010). *Peer Review: A guide for researchers.*
- Technopolis Group. (2010). *REF Research Impact Pilot Exercise Lessons-Learned Project: Feedback on Pilot Submissions.*
- Van Raan, Anthony, F.J. (2005). *Challenges in ranking universities.* First International Conference on World Class Universities. Shanghai.

About LERU

LERU was founded in 2002 as an association of research-intensive universities sharing the values of high-quality teaching in an environment of internationally competitive research. The League is committed to: education through an awareness of the frontiers of human understanding; the creation of new knowledge through basic research, which is the ultimate source of innovation in society; the promotion of research across a broad front, which creates a unique capacity to reconfigure activities in response to new opportunities and problems. The purpose of the League is to advocate these values, to influence policy in Europe and to develop best practice through mutual exchange of experience.

LERU publications

LERU publishes its views on research and higher education in several types of publications, including position papers, advice papers, briefing papers and notes.

Position papers make high-level policy statements on a wide range of research and higher education issues. Looking across the horizon, they provide sharp and thought-provoking analyses on matters that are of interest not only to universities, but also to policy makers, governments, businesses and to society at large.

LERU publications are freely available in print and online at www.leru.org.

LEAGUE OF EUROPEAN RESEARCH UNIVERSITIES

Universiteit van Amsterdam • Universitat de Barcelona • University of Cambridge • University of Edinburgh
Albert-Ludwigs-Universität Freiburg • Université de Genève • Universität Heidelberg
Helsingin yliopisto (University of Helsinki) • Universiteit Leiden • KU Leuven • Imperial College London
University College London • Lunds universitet • Università degli Studi di Milano • Ludwig-Maximilians-Universität München
University of Oxford • Université Pierre et Marie Curie, Paris • Université Paris-Sud 11 • Université de Strasbourg
Universiteit Utrecht • Universität Zürich

LERU Office
Huis Bethlehem
Schapenstraat 34
B-3000 Leuven
Belgium

tel +32 16 32 99 71
fax +32 16 32 99 68

www.leru.org
info@leru.org