

Towards making research evaluation more compatible with developmental goals

Rodrigo Arocena¹, Bo Göransson² and Judith Sutz^{3,*}

¹Faculty of Sciences, University of the Republic, Aiguá 4225, PC 11400, Montevideo, Uruguay, ²School of Economics and Business Administration, University of Lund, Lund, Sweden and ³University of the Republic, University Research Council, Jackson 1303, PC 11200, Montevideo, Uruguay

*Corresponding author. E-mail: jsutz@csic.edu.uy

Abstract

Research evaluation practices linked to social impact have important systemic effects on the prioritization and organization of research while at the same time leading to the delivery of higher social value. Amidst growing criticisms, global research evaluation has evolved in a different direction, characterized by quantitative metrics and mimetic behavior. The article deals with the forces that sustain the prevailing research evaluation system, asks why it has proven to be so resilient, and discusses alternative proposals. A new argument for building an alternative is put forward: the need for a developmental role for universities, introducing the notion of ‘connected autonomy’ allowing universities to productively and in a nonsubordinated way collaborate with a broad set of actors to achieve desirable social changes. An outline is presented for how to make research evaluation practices and the pursuit of developmental goals more compatible, an important issue for knowledge public policy.

Key words: research evaluation; developmental universities; connected autonomy; knowledge democratization; devmetrics.

1. Introductory remarks

This article shares the concerns of a growing body of literature regarding the inadequacy of current practices of research evaluation to deliver stated goals such as advancing science and benefiting society through the creation of new knowledge. The article deals with these concerns from a developmental perspective, justified because some developmental problems today affect the whole world; moreover, research evaluation has become an internationalized practice with striking similar consequences, all socio-geographical differences notwithstanding.

Development is not any more an issue involving backward countries struggling to catch-up with the developed countries. ‘Human development’ and ‘sustainable development’ have widened the concept in scope while also narrowing the number of countries that may be considered developed if these facets are fully considered. There are rich countries (by some common indicators) and poor countries. Aiming at becoming richer, some of the latter started following a path of rapid industrialization while other pursue strategies of intensive short-term extractive activities, adding stress on the environment, even though the rich countries continue to be a leading cause of its long-term degradation (Milanović 2016). The danger of irreversible damage is grave, urgent, and global. The United Nations Sustainable Development Goals—SDG—(UN 2015) reflect this, pointing specifically to two issues that presently ravage societies in North and South: environmental problems and inequality.

The SDG pose concrete challenges for which answers are expected a bit more than a decade ahead. Research has a say in the building of these answers. To live up to the expectations from a developmental standpoint, research will need to address new problems—not necessary problems that are new in themselves, but that have so far largely been excluded in research agendas—and find new types of answers. Both of these aspects are equally important. Vast parts of the world population belong to the ‘have nots’ in terms of goods and services that can be considered basic for a decent living and are available for the rich in poor countries and for almost everyone in rich countries. The extent of resources needed to expand their reach to the required level, at current prices, is overwhelming. Alternative ways of looking for answers—because a given problem never has one and only one good solution—along with new questions, requires the shaping of a huge—and rather new—research program. New names have been given to approaches able to face such challenges, such as responsible research and innovation (RRI) (European Commission 2012), with emphasis on direct interactions with society.

There is room for skepticism regarding the possibilities of implementing this new research program and the advancement of the SDG in general. But the sense of urgency at a global level opens spaces for action. New questions arise: which institutions need to be involved in the implementation in such a research program, what changes should they undergo to be able to play the new game, how can at least some of its main actors get involved with those changes?

University is clearly one of the institutions to be enrolled, because it is a fundamental site of knowledge production and because it is the place where young people should become aware of their social responsibility in putting knowledge at the direct service of a human and sustainable development.

Related commitments need to be built proactively and the obstacles that hamper its expression need to be overcome. Both issues point to incentives, to what is rewarded and what is discouraged in academic life. Research evaluation is today an important obstacle to the redirection of academic commitments towards the advancement of the SDG. Research evaluation is the outcome of a complex social process where shifts in power over academic life are apparent (Gläser 2010); designing feasible alternatives thus needs to consider power relations.

The article aims at making a modest contribution to this. But before sketching a feasible alternative to research evaluation practices, it is important to analyze the prevailing research evaluation system, why alternatives to it are needed if fostering the SDG is considered part of the goals of the research effort, what are the main flaws of the prevailing evaluation system regarding the latter, why changing the system has been so difficult, and what is the spirit of the proposed alternatives so far. This analysis is done through a thorough revision of the literature. Methodologically, the path through the literature gives three levers to the article: first, the saturation point regarding the critics to the prevailing research evaluation, involving North and South, shows a world spread phenomenon; second, the proposed alternatives show several traits in common, showing again a sort of convergence in what is considered desirable; third, the resonance of the findings and proposals appearing in the literature with the authors' experience of problems in their own universities suggests that the alternative practices to research evaluation sketched in the article may be of interest to an increasing audience.

Many voices claim that the research effort should take on board public-engagement, inclusive participatory multi-actor dialogues, co-creation of research and innovation, and citizen-centric action research.¹ For that, the article discusses the concept of 'connected autonomy' of universities as an important trait that needs to be cultivated, which has consequences for research evaluation. The sketch of an additional alternative to research evaluation is then proposed, based on the lessons stemming from the presented analysis.

The article is organized in the following six sections, dealing respectively with: (Section 2) some of the consequences of the prevailing research evaluation system; (Section 3) main reasons behind the present situation and its resilience in spite of the increasing concerns it raises; (Section 4) critiques of the prevailing research systems and alternative proposals; (Section 5) the rationale for an additional alternative proposal linked to the concept of 'connected autonomy' of universities, and (Section 6) a sketch of such proposal. A brief concluding Section (7) reflects on the feasibility of the proposal.

2. Where is the prevailing research evaluation leading us?

The current practices of research evaluation can be characterized by the opening phrases of the Leiden Manifesto (Hicks et al. 2015: 429, emphasis added): 'Data are increasingly used to govern science. (...) *The problem is that evaluation is now led by the data rather than by judgement.* Metrics have proliferated: usually well intentioned, not always well informed, often ill applied.' Analyzing the effects of these practices offers answers to the question posed in the title of this section.

After being in place for 20 years, the evaluations implemented by the Mexican National System of Researchers (NSR), heavily based on publication counts have been found to '...act against the activities more directly related to institutional objectives different from research. (...)'; and, consequently, '...the institutional diversity is being reduced, with serious consequences for the integral goals and the joint productivity of the system.' The NSR 'is inhibiting the quality and the creativity of researchers (...). Researchers simply comply with the requisites of the NSR. Researchers do not want to enter into much riskier projects in which they may pass 3, 4, or 5 years without publishing anything because the project they get involved with is too complex and will not allow to have in time the scientific publication required by the NSR.' (Foro Consultivo Científico y Tecnológico y Academia Nacional Mexicana de Ciencias 2005: 50, 54, our translation) A recent account of the consequences of the Mexican NSR for ecological research confirms the former assertions (Neff 2017).

The British Research Assessment Exercise (RAE), also based on publications counts, is said to be '...reductionist and primitive, and almost certainly counterproductive in terms of generating a wide variety of intellectual innovations in the longer term.' (Martin and Whitley 2010: 75) Reviewing evidence going back 25 years, the same authors observe that the RAE seems to have induced a bias leading to basic rather than applied research, shorter-term rather longer-term research, incremental rather than more ambitious or 'pioneering' open-ended research, monodisciplinary rather than inter- or multidisciplinary research, 'academic' rather than 'professional' research (for instance, in medicine, management, law, planning), research that yields journal articles rather than books (ibid: 70). It has reduced '...the willingness of faculty to engage in other academic activities such as reviewing, editing, translating, contributing to reference works, writing popular books, engaging in clinical medicine or community service, providing policy advice, and so on...' Concerning personal life, it 'has been a factor encouraging overwork and adding to levels of stress. It has disadvantaged those (predominantly women) who have taken time off for family or other reasons, resulting in a 'gap' in their published output.' (ibid: 72).

Particularly worrisome are the effects of research evaluation on young researchers. 'Early career researchers can feel particular pressure to perform well against certain indicators' (Wilsdon et al. 2015: 93); this negative side-effect has also been observed in Sweden and Mexico (Hammarfelt and de Rijcke 2015; Neff 2017).

Research evaluation influences university practices. The direction of such influence is not clear, though. In terms of ideal types, the 'Entrepreneurial' university and the 'Humboldtian' or Academic university are usually considered. Other models emphasize the direct relationship with society as the civic universities (Goddard 2009) or the developmental universities (Arocena et al. 2018). The entrepreneurial university, believed to provide economic salvation (Halfman and Radder 2015: 172), has become the leading model for transforming the university worldwide. The prevailing academic evaluation system, however, seems to point in another direction. The British RAE, for instance, 'has largely enhanced the authority of established scientific elites rather than orienting academic research towards economic goals and what the state considered to be user needs' (Whitley 2010: 37). Current evaluation patterns 'reproduce classical conceptions of knowledge and power relations. They encourage a return to 'ivory tower' research conducted by elites in selected institutions at a time when complex global problems and policy objectives require the involvement of interdisciplinary teams

with diverse perspectives and experiences' (Hazelkorn 2009: 11). Rather than promoting the Entrepreneurial University, the actual result seems to reinforce the Academic University. Nevertheless, it is not obvious that the prevailing criteria are helpful even from the point of view of the Academic University: 'If researchers only focus on delivering short-term accountable results and managing their publication assets, what will happen with the long-term basic research that may deliver results that are important in 20–30 years?' (Kronman 2013: 124) For some scholars, that future has already arrived: 'The numbers of journal articles published by a researcher and the level of the journal in which they appear has moved from a modest issue to a major concern. For some it has become almost the only concern. Having something important, relevant, and meaningful to say seems to have become comparatively less important than doing and publishing research that appears in the right journal' (Alvesson and Spicer 2016: 32).

The current evaluation system does not either promote a 'socially engaged' university. 'Long-term involvement with society, which is both complex and uncertain, sits at odds with an academic career progression that values a constant stream of research outputs.' (Trencher et al. 2013: 20) Moreover '...the effort required to engage with civil society and to use societal challenges as the fundamental platforms for much of the work can actually distract from the raw pursuit of academic excellence.' (Brink and Hogan 2016: 252–53)

Research evaluation has gone global. Homogenization of consequences has accompanied homogenization of practices. In the case of developing countries, homogenization implies complying with the rules, following a mimetic path in terms of what constitutes a good researcher, with its set of inducements resulting from standardized measurements. 'The Third World looks to the North for validation of academic quality and respectability. For example, academics are expected to publish in Northern academic journals in their disciplines. Promotion often depends on such publication. Even where local scholarly publications exist, they are not respected.' (Altbach 2003: 6) Instead of concerns for the impacts that quantitative measures of research productivity may have on young scholars, we mainly found praise for its disciplining effects. This is the case in Uruguay: 'Researchers, especially the youngest that represent the entrance door to the National System of Researchers, want to respond to the requirements of the instrument, which gives it an important normative power' (National Agency on Research and Innovation, ANII 2012: 7, our translation). This is not the whole story, though: as early as 2004, young researchers from the exact and natural sciences in Uruguay criticized the use of journal rankings to evaluate individual performance and proposed alternatives, very much in tune with what several years after would become common understanding among scholars around the world. (PEDECIBA 2004) But their warnings and suggestions were not enough to influence academic practices: until today, young researchers belonging to the National System of Researchers in Uruguay receive in each evaluation exercise the same type of recommendation: 'you should be more productive and publish in high impact journals.'

Mimetic behavior includes using similar metrics and aiming at similar results, biasing thus the choice of research problems and methods. This is not the genuine international collaboration among researchers from different countries, but the subtle conditioning stemming from the perceived thematic preferences of main stream journals. Working on problems that are specifically tied to developing countries is not a good recipe for acceptance in mainstream journals: 'Imposing an evaluation system that privileges international

citations will force scholars to choose topics that interest foreign academics...' (Hicks 2013: 79). Where research is weak, this trend further weakens its capacity to address problems of local relevance (Bianco et al. 2016). From the SDG perspective, this is perhaps the most dangerous outcome of the current research evaluation system.

3. Why has research evaluation evolved towards the prevailing situation and why is it so resilient amidst growing criticisms?

The forces that drove the research evaluation to its current state as well as those that support its continuation will need to be counter-acted to some extent in order to make room for alternatives; to find at least an approximation to the questions posed above is thus important.

Research evaluation has dramatically changed in recent times, both in scope and in methods. There are powerful currents 'whipping up the metric tide', including 'growing pressures for audit and evaluation of public spending on higher education and research; demands by policymakers for more strategic intelligence on research quality and impact; the need for institutions to manage and develop their strategies for research; competition within and between institutions for prestige, students, staff and resources, and increases in the availability of real-time "big data" on research uptake, and the capacity of tools for analysing them.' (Wilsdon et al. 2015: viii, emphasis added) These processes have accompanied an enormous growth of advanced knowledge and Higher Education, as well as its qualitative diversification. Research evaluation is thus increasingly perceived as unmanageably complex and time-consuming. The search for simplicity has fostered a great shift in evaluation that favors quantitative methods and standardized tools that can be used across different areas (Dahler-Larsen 2013). A Committee appointed to propose a system of quality indicators for the humanities in The Netherlands states: 'The Committee finds it important to note that the expectations of some policy-makers are too positive regarding a 'simple' and preferably fully quantified system that makes easy comparisons possible between research groups and even whole disciplines' (KNAW 2011: 10). The push towards simplification is thus part of the answer to the first part of the question posed in this section.

The process of marketization of universities in some highly industrialized countries also plays a role. The services they sell, particularly to foreign students, have a noticeable impact in the economy. In the case of the UK, international students spent in 2014–15 'an amount equivalent to 2.6% of all UK exports of goods and services in the same year'. (Universities UK 2017: 3) Competition in the HE market has become fierce, leading to a 'gladiator obsession' with international rankings, where '...increasing credibility derives from their simplicity and provision of 'consumer-type' information independent of the HE sector.' (Hazelkorn 2007: 1) The present situation has been characterized as '...the paradox of increased rationality breeding the irrational cult of stardom' (Magala 2009: 133). For many highly industrialized countries, 'The ultimate end of the benchmarking, ranking and scoring is to improve the commercially exploitable position of an institution or a part of an institution among other suppliers and to become preferred suppliers of knowledge products to an increasingly diversified variety of clients' (ibid.). Universities in developing countries can hardly be considered international suppliers of knowledge products. Nevertheless, the

poor research performance exhibited by comparative rankings lead academic elites in these countries to further pursue the academic mimetic behavior already mentioned.

Money, simplicity, and analytical strengths to handle data are certainly not the only explanatory factors of the present situation. ‘Metrics hold real power: they are constitutive of values, identities and livelihoods’ (Wilsdon et al. 2015: iii). This constitutive capacity implies that along the way new factors may emerge which reinforce current trends. One such factor is the self-preservation interest of managerial staff involved in research decision-making whose number and power have exponentially increased over the last decades.

Coming to the second question—why does it appear to be so difficult to change the research evaluation system if so many observers consider it inappropriate?—the issue is to identify actors or situations able to exert such power over the system as to shield it against attempts to change it. Tentatively, three such situations may be suggested. Perhaps the prevailing evaluation system persists because a sort of implicit agreement exists that keeps the system working: governments and academic institutions can show an example of their commitment to New Public Management (NPM) while academic elites are satisfied with what really is a means of enhancing their world-view and power. In fact, research management cannot act without previous advice on what counts as quality research, which is provided by the academic elite. This objective conjunction of interests protects evaluation patterns that seem, at least in the long run, academically damaging. A second situation refers to a sort of gagging induced by the fundamental assertion of current evaluation that there is a strong correlation between bibliometric scores and academic reputation. Therefore ‘...researchers with a good reputation have nothing to fear and have no particular reason to be critical. The breathtaking implication, never articulated, is of course that the critical academic was critical because he was not a good researcher’ (Dahler-Larsen 2015: 22). If the first situation refers to a ‘policy power’ over researchers, the second one refers to a related ‘subjective power’, which has a distinctive expression in researchers’ feelings and behaviors. This kind of subjective power also leads to the third situation that can be suggested to explain the paradox. Young scholars, even if they are strong critics of the system, feel defeated by the research evaluation system, because before they may raise their voices against it, they need to build academic legitimacy. This implies a long period of putting most of their energies to address the systems’ requirements, because the dominant feeling is that ‘we are the journals in which we have published in’ (Alvesson and Spicer 2016: 39). This way of understanding ‘who we are’ cannot be kept in a modest undertone. The current evaluation system belongs to the category ‘glass-cage organization’ (Gabriel 2005), ‘characterized by its constant visibility and incessant surveillance’ (Alvesson and Spicer 2016: 40), which is supported by the technical possibilities offered by immediate and online information on publications. It is thus understandable that young scholars, who would gain more from being freed from the glass-cage, feel too weak to challenge the system. The combination of policy power and its derivative subjective power present a formidable barrier to the emergence of alternative practices of research evaluation.

4. The nucleus of criticisms of current research evaluation and alternative proposals

Criticism of current research evaluation has come near to what in sociology qualitative research is termed ‘saturation’: mostly the

same arguments appear in the increasing number of articles related with the general problems present in today’s research evaluation practices. One of such arguments is the performative capacity of research evaluation, which may lead to unintended consequences, such as giving a bias to research topics: ‘Researchers respond to priorities set by funding councils and to their perception of peer review being biased towards mainstream and low-risk research’ (Gläser and Laudel 2016: 125). Another problem is the constitutive capacity of the research evaluation system to lead to errors, due to what is termed Goodheart’s law, formulated on STI matters as: ‘once STI indicators are made targets for STI policy, such indicators lose most of the information content that qualify them to play such a role’ (Freeman and Soete 2009: 583). Two almost unanimously reported problems are goal displacement (Butler 2007), which occurs when the measure itself becomes the goal instead of a means to appreciate the achievement of a goal, and the hindering of interdisciplinary research.

Criticisms come in different formats. Some are found in recent ‘manifestos’, notably DORA (Declaration on Research Assessment) 2012, and the Leiden manifesto 2015. They provide a combination of critiques and alternative proposals which purpose is to influence particularly the quantitative part of research evaluation: ‘We offer this distillation of best practice in metrics-based research assessment so that researchers can hold evaluators to account, and evaluators can hold their indicators to account’ (Hicks et al. 2015: 430). Other sources of criticisms come from country analyses, from reviews commanded by national authorities and from supranational organizations like the OECD (Council of Canadian Academies 2012; Wilsdon et al. 2015; Zaccai et al. 2016, OECD 2009).

But the issue has not been settled. Saturation appears among the critics but expectedly those who explicitly praise the rationale of the prevailing evaluation system exist and are powerful. For example, on a sensitive issue, performance-based funding, while the European Commission recommends increasing ‘the share of performance-based funding of research institutions’ (Jonkers and Zacharewicz 2016: 9), a panel organized by the Canadian Academies of Science states a different position: ‘...science indicators—essentially measures of past performance—may not provide a reliable guide to future prospects. Overall, the Panel found no evidence that there is a single correct funding response to any assessment results’ (Council of Canadian Academies 2012: xiv) Regarding the impact of research evaluation on national science, opposite assessments can be found; the Mexican NSR illustrates well this (Neff 2017; Williams and Morrone 2018).

Alternative proposals and recommendations usually accompany criticisms. Inspired by the approach to Responsible Research and Innovation mentioned above, the concept Responsible Metrics has been proposed (Wilsdon et al. 2015: 134). Its defining features are robustness, humility, transparency, diversity, and reflexivity (ibid.). Humility implies recognizing the support or complementary role of quantitative indicators in research evaluation; diversity, in a nutshell, implies recognizing that one size does not fit all across institutions, disciplines or research maturity; reflexivity implies acknowledging that the use of indicators may have effects beyond the intended ones, paying attention to their appearance and changing indicators accordingly. Similar recommendations have been forcefully put forward in several other texts (Butler 2007; Council of Canadian Academies 2012). The Canadian document presents four guidelines for research evaluation; one is transparency. The other three are: context matters, stressing that research assessment strategies depend on the objectives, goals and priorities of the

organizations where evaluation is implemented; do not harm, akin to the former reflexivity, alerts to the potential of unattended consequences, and recommends humility to face it, that is, a balanced set of indicators and expert judgement, that remains invaluable, this being the fourth guideline. Their judiciousness notwithstanding, these recommendations—as well as those of the manifestos—do not go in line either with simplicity or with the compliance to the powers that be. Its implementation is not only a matter of offering convincing arguments but of building alliances of stakeholders able to stand up for them: it cannot come as a surprise that adoption of the alternative proposals and recommendations is not advancing at a rapid pace.

For the SDG, the implementation of these recommendations would be of great help. This would allow for a ‘re-centering’ of priorities where complex problems requiring long terms efforts and interdisciplinary approaches may be thoroughly addressed without fear of reprisal from evaluation. A further step is needed here: research aiming to contribute to development needs to be assessed in terms of the likelihood of the impact it promises to achieve. Recalling the performative effect of evaluation criteria, special care is needed to devise indicators that give visibility and value to activities without which research would remain at arms-length from society. Linking research evaluation better to social impact can have important systemic effects on the prioritization and organization of research while at the same time leading to the delivery of higher social value for the same amount of funding: ‘Vast improvements in the scientific system could be had if science agencies strengthened the ties that link research agendas to social needs, and counteract the perverse incentives that commit scientists to careers measured by publications and grant dollars rather than the creation of socially valuable knowledge’ (Sarewitz 2017: 139).

Steps in that direction have been recently taken in some places. In the UK: ‘Following considerable debate, and an extensive pilot exercise carried out in 2010 [...] the impact assessment was made by a process of expert review, with users of research from outside of academia playing an important role. In contrast with the assessment of outputs, no systematic gathering or use of quantitative data on impact was carried out to support the assessment’ (Wilsdon et al. 2015: 126-27). Societal impact—even if with diverse meanings—has been also included, for instance, in the NSF Broader Impacts Review Criterion and in the Canadian Institute of Health Research.

Neither the importance of this move nor its associated difficulties should be underestimated. Social impacts are difficult to highlight because ‘...science values and economic values are available as accepted and dominant surrogates for all other values’ (Bozeman and Sarewitz 2011: 7). But: ‘... why begin with surrogate values? Is it not more sensible to premise policies on the outcomes they should achieve rather than the instruments presumed (perhaps erroneously) to enable those outcomes?’ (op.cit.; 6).

Research aiming directly at socially desired outcomes should be assessed in terms of social impact, in combination with conventional criteria. This is easier said than done; ‘...there has been remarkably little progress in the ability to measure directly, systematically, and validly the impacts of research on social change’ (Bozeman and Sarewitz 2011: 3). The complexity of the endeavor plus the lack of attention to the whole issue so far explains the exploratory nature of the attempts to build an alternative proposal in this realm of research evaluation.

There is an additional point worth mentioning. Research evaluation accentuates what are deemed worthwhile efforts; what is not valued is rendered invisible. If part of the university research wants

to be oriented directly to social problems of the type included in the SDG, signals need to be given that such orientation would be valued. Research geared directly towards social issues requires a process of problems’ identification and characterization that differs from that of purely academic research and from the commanded research of external actors. Such a process is part of a wider movement of university embeddedness with society that needs also to be considered in the research evaluation alternatives. We call such a movement-connected autonomy and we now turn to its consideration.

5. An additional argument for finding alternatives to current research evaluation: SDG need developmental universities endowed with connected autonomy

Universities are evidently important institutions for research oriented toward the advancement to the SDG, as knowledge producers and educational spaces for young researchers and professional knowledge users.

To produce socially valuable knowledge, universities should not be subordinated to outside forces. Autonomy in this sense means freedom to pursue research in directions that are internally decided, utilizing appropriate methodologies, and communicating the obtained results regardless of who may find them inconvenient to their interests. But university autonomy is by no means fully defined by these features. It may be fulfilled while ignoring the society in which the university is inserted, claiming to perform high-quality research defined exclusively in their own terms. Alternatively, autonomy may be conceived in such a way that, along the fulfillment of its defining traits, its aims include to better serve society by considering voices outside academia: ‘... working on the supply side of the knowledge economy is necessary and valuable but not sufficient. That is why the University seeks also to work on the demand side. Such work, or ‘engagement’, is responsive in nature, and is determined not only by the curiosity of the researcher, but also by issues and opportunities arising from within and across global society.’ (Brink and Hogan 2016: 240) The first way of conceptualizing autonomy, the ‘ivory tower’, is akin to autarchy; the second one, that implies recognizing the existence of legitimate ‘external’ university stakeholders and fosters dialogues with them, may be named ‘connected autonomy’ (Arocena 2015).

Universities’ ‘connected autonomy’ presents a resemblance with states’ ‘embedded autonomy’. (Evans 1995) Although there are obvious differences, mainly related to issues of power, the two concepts point to similar dynamics.

Autonomy needs to be shielded from the possibility of institutional usurpation by stakeholders with power to impose policies, behaviors or ideologies to their advantage: this holds, even if in different forms, for universities as for states. But shielding may be as well asphyxiating, leading to an institution that only serves itself. ‘The problem is separating the benefits of insulation from the costs of isolation.’ (Evans 1995: 41) To avoid isolation, some ways for immersing in society are needed; strong voices against that arise. For states, it is said that market-friendly policies are enough to achieve optimal economic outputs while industrial polices can only lead to committing costly mistakes. For universities, some posit that the pursuit of knowledge for its own sake maximizes the harvest to be put at social disposal while deviating from this path harms the quality of the scientific enterprise. Balanced outcomes are always

difficult to achieve, and they remain unstable amid changing circumstances. It is important not to reify institutions by ascribing to them volition of their own. 'In practice "the state wants" because some group of individuals within the state apparatus has a project. (...) their project may well be opposed by others elsewhere in the state and (...) the definition of what the state "wants" is the result of internal political conflict and flux.' (Evans 1995: 19) What 'universities want' is the outcome of a complex web of interactions involving different internal groups, the governance system, the financial regime, the academic system of incentives, and the kind and strength of the university relations with the rest of society.

Connectedness is the outcome of considering the knowledge needs of different social actors when producing and sharing knowledge. In democratic and pluralistic societies, universities should be connected to the whole gamut of social actors. This clearly is not the case. Hess (2007) refers to the efforts that are lacking to help building the necessary knowledge to tackle social problems as 'undone science'. The concept refers to social actors that need knowledge support for their aims and concerns and cannot find it: this is an expression of lack of connectedness. A more complicated situation appears when social actors that would have benefited from a specific cognitive approach to their problems are not able to identify research as a tool for solving them: the most deprived part of the population is usually in this situation. Other actors can have their knowledge demands taken on board, like big business firms in knowledge-intensive sectors and sections of the state. Such actors are often the drivers of the connection: in exchange for the services they need, they offer material or symbolic resources that the university requires for its functioning.

Two similarities between the embeddedness of the state and the connectedness of the university are worth mentioning. First, both state and universities need to take the initiative in building the conditions to establish the linkages and setting their rules. Second, this building needs autonomy to avoid been captured by particularistic interests.

Evans states (1995: 245) that 'capacity without connection will not do the job'. This holds as well for universities, if the job is understood as maximizing the social usefulness of the knowledge produced, taught, and disseminated. To what extent universities become socially connected depends both on internal struggles and on how the innovation system in which they are inserted behaves; differently from states, that in Evans' parlance may 'build' the actors with which they become embedded, universities have to a much lesser extent the power to do that.

Developmental research agendas require a rich variety of stakeholders with whom they together build the problems they will contain. This is related with the idea of co-production (Ostrom 1996) and with considering people as agents and not as patients (Sen 1999). Developmental research agendas depend crucially on the level of connectedness of universities. Universities that include fostering such agendas may be named 'developmental universities', characterized by the joint promotion of teaching, research and extension activities in cooperation with several actors with the overall purpose of contributing to human and sustainable development, with special attention given to the democratization of knowledge (Arocena et al. 2018).

If (i) the university is not aware of the demands or needs of a social actor, (ii) this actor does not have the capacity to pay for research on his problems or interests, and (iii) he is not able to take initiative to become connected to universities, then the university needs to take a pro-active stance to avoid that those actors remain

invisible for research agendas. If connectedness is something that occurs only when an external actor wants to get in touch, the concept is narrowed in practice. If specific challenges faced by marginalized actors are to be considered in research agendas, the complex process of dialoguing, searching, understanding, translating and co-constructing problems that this entails needs explicit support from the university. Such support is indeed a constitutive trait of an engaged university: 'Engagement has to be an institution-wide commitment, not confined to individual academics or projects. It must embrace teaching as well as research, students as well as academics, and the full range of support services' (Goddard 2009: 4).

The social involvement of student and faculty requires receiving formal academic status; both need to feel backed-up for devoting time and efforts to issues that may take more time than average to show progress because, among other reasons, they involve non-academic actors. Universities need to accommodate the academic incentive system to legitimate related activities and '...at least take the obstacles away for those who like to work on societal impact in the academic setting, and facilitate their work.' (Borg et al. 2016: 216) Connectedness needs indeed to be facilitated, involving additional alternatives to current research evaluation.

6. A sketch of an additional alternative to research evaluation

Research agendas are built around and influenced by diverse interests and incentives. Academic evaluation is a particularly powerful influence: prevailing criteria, unilaterally emphasizing academic status, hamper more than promoting the developmental role of universities. New, more balanced, pluralistic and 'developmental' evaluation systems are needed in order to foster both academic quality and social engagement. As a contribution in that direction, in this section we sketch first a 'devmetrics' and then some characteristics of an evaluation system that includes such metrics but is not restricted to it.

The 'devmetrics' denomination may appear misleading at first sight. To start with: is devmetrics a metric at all? It attempts at measuring some features, but not at measuring how well recognized a piece of academic work is. In this sense, it differs from current bibliometrics, from broader bibliometrics—that includes a wider set of journals and datasets—and from altmetrics as it is currently defined, with its emphasis in capturing all the ways in which web-based references may highlight a given academic contribution, including non-academic references. However, altmetrics can help devmetrics by unearthing hidden references that indicate social use of results; moreover, devmetrics may in the future belong to altmetrics, given its alternative character and its aim at fostering and recognizing differently the worth of research work.

What does devmetrics attempt to measure? Results and efforts: As it is argued below, some type of results needs specific efforts from researchers to build, with other social actors, the research that may lead to those results: if they are not recognized and measured, such efforts will not be deployed. In this sense, devmetrics recognizes that research evaluation deals with past actions but influences as well future actions, and so evaluating what has not yet been achieved but paves the way to future achievements should be part of research evaluation efforts.

A 'devmetrics' is defined by the purpose of putting knowledge at the service of sustainable human development, with priority given to deprived populations. What would it look like? It would share several of the alternative emphases proposed so far: attention to

context, robustness, humility, transparency, diversity, and reflexivity. It would be related to the aim of universities. It would take care not to discourage interdisciplinary work, given that this is needed to deal with the demanding problems societies are currently experiencing.

Before suggesting issues, methods and tools related to such criteria, it must be acknowledged that the task of devising a ‘devmetrics’ is far from straightforward. Several difficulties may be pointed out. Operationalizing normative aims in research is different from stating how many children should attain a full vaccinating scheme in a determined period. Developmental aims usually cannot straightforwardly be related to research agendas and research outputs. The difficulties of defining and applying comparable criteria to assess university–society collaborations have been amply acknowledged (e.g., Bölling and Eriksson 2016; Molas-Gallart and Castro-Martínez 2007). Research results impact the fulfillment of the envisaged normative aims through mediations that are outside academia. The latter notwithstanding, the orientation of research does play a role in such fulfillment. How can a metric appreciate this?

In developing countries, research systems are weak as a rule by almost all usual indicators; making science there is particularly difficult. However, a developmental university must perform first-rate research to promote creativity among its students and to address with some success the challenging problems that affect the society to which they belong. Such problems, as remarked in the critical literature on research evaluation, probably will not raise much interest in the journals supposedly assuring first-rate quality. Evaluation strictness as if research was context-free would further foster individualist behaviour and disregard of local problems; evaluation permissiveness just because research is geared towards local problems would do little to enhance the capacity of local science to solve them. How can a metric consider the context in which research is done without fostering a culture of isolation that would do little more than self-justifying mediocre research?

To approach these questions, it is worth asking another one: which are the strongest constraints that prevailing evaluation systems put on ‘developmental’ research? Basically, pressures of time and prestige. Countervailing these constraints cannot—and should not—be done by throwing away concerns about research productivity and international visibility. But productivity should be understood differently from counting papers published in a given set of journals. In particular, researchers that have proven their capabilities by means of a serious contribution duly recognized by the academic community should be awarded the freedom to tackle very difficult problems that probably would not lead to a quick harvest of papers. The evaluation criteria in these cases should be an assessment of the difficulty of the problems and the soundness of the strategies to deal with them.

Doing developmental research involves devoting time to determining the research problem and to assuring the fulfilment of the correct methodological conditions for research itself. Engaging in co-production of socially relevant research with non-academic and excluded parts of society will be exploratory by nature and, thus, quite time-consuming in both the needs identification as well as the design of the method for tackling the problems. Evaluation should reflect these particular circumstances and consider not only the outputs and outcomes of research but the inputs to research that the researcher must construct.

The prestige issue is more complicated to deal with. Building a counter-hegemonic prestige regime is fundamental, even if only aiming at complementing the current one. This is not principally a question of rewarding financially the fulfilment of a different set of

criteria. ‘Although the importance of (performance-based research funding systems) PRFSs seems based on their distribution of universities’ research funding, this is something of an illusion, and the literature agrees that it is the competition for prestige created by a PRSF that creates powerful incentives within university systems’ (Hicks 2012: 251). Even if in a different realm, an alternative proposed to the patent system in health resonates forcefully: ‘There is an alternative way of financing and incentivizing research that, at least in some instances, could do a far better job than patents, both in directing innovation and ensuring that the benefits of that knowledge are enjoyed as widely as possible: a medical prize fund that would reward those who discover cures and vaccines. (...) That said, *the prize fund would not replace patents. It would be part of the portfolio of methods for encouraging and supporting research*’ (Stiglitz 2007: 46–7, emphasis added) Prizes can be local, national or even international, they may or not be richly endowed, but if the awarding committee is prestigious enough, they could be a helpful tool for building an alternative to the current way of assigning research prestige, particularly so in developing countries.

The above considerations suggest that a devmetrics should:

- a. Give particular relevance to the issue of fostering connectedness, including co-construction of problems and interactive approaches to research strategies and implementation of results.
- b. Use mainly qualitative evaluation methods, particularly narratives and interviews with external actors.
- c. Concerning evaluation tools, during the period to be evaluated the researcher could be required (i) to show her capacity to do good research by having at least one publication in a good journal, but neither the number of publications nor the place of the related journals in the journals’ prestige ranking will be considered; (ii) to give detailed account of the strategy she followed to build a research project with external stakeholders that takes their needs into account, the results achieved in research terms—including formal papers as well as ‘grey’ knowledge products—and the results achieved in developmental terms. The latter is difficult to portray; it can only be approached through preliminary results and subjective appreciations. Knowing this beforehand, and so not being fooled by any illusory objectivity, may help to perform a reasoned and convincing research evaluation exercise.

Evaluation has several levels—individual researchers, departments, whole institutions—and criteria may be different in each of them. To encompass to some extent the former considerations in the case of individual researchers, a pluralistic evaluation system would need to be conceptualized as a tool ‘to improve or shape the research contents’ (Råfols et al. 2016: 3). Three main guidelines for that may be proposed.

- i. Allow for flexible shifts from prevailing metrics to ‘devmetrics’: researchers may work for a period being evaluated by one metrics and for another period being evaluated by the other, depending on the working program they commit themselves to pursue;
- ii. Value the time devoted to help widening the cognitive connectedness of the university as research time;
- iii. Offer prizes for the best research outcomes that help solving societal problems; devise specific calls for research aimed at finding this type of solutions.

An attempt to summarize the guidelines for such an alternative evaluation system is presented in Table 1.

Table 1. Identifying characteristics of an evaluation system that includes a 'devmetrics'

Issues	Characteristics
Aims of research	Contribute directly to addressing social problems, particularly those affecting vulnerable populations; contribute to the meeting of the SDG
Modes of evaluation	Mixed: canonical and 'devmetrics', successive or integrated
Modes of inducement	Prizes for research outcomes that help solving societal problems
Issues to be valued in the 'devmetrics' mode (ex-ante)	Efforts towards connectedness. Objective results of such efforts, for instance: (i) links created with public policies and collective actors; (ii) reformulation of social problems as research questions; (iii) gathering of the necessary contributions, presumably interdisciplinary, to generate the knowledge able to tackle such research questions; (iv) advances, even if incipient, in the generation of such knowledge
Method of valuation in the 'devmetrics' mode	Mainly qualitative: narratives, interviews with external actors
Evaluation tools in the 'devmetrics' mode (ex-post)	During the period to be evaluated the researcher is required: (i) to show her capacity to do good research; (ii) to give detailed account of (a) the strategy she followed to build a research project with external stakeholders that takes their needs into account, (b) the results achieved in research terms and (c) the results achieved in developmental terms
Time-framework of evaluation	The evaluation period for individual researchers' performance allows for the time needed to build research problems with external stakeholders

The implementation of the principles of 'devmetrics' in the evaluation system would constitute a clear break away from its present constrained form, towards broader social engagement where developmental research is better recognized and appreciated. Obviously, the guidelines would have to be operationalized in the form of concrete actions and the specification of 'devmetrics'-based evaluation tools. Such an operationalization will no doubt be context dependent in the sense that each attempt to change will face rather unique power constellations necessitating the development of customized approaches in the implementation. A start would be for the leaderships of research councils and universities to clearly state the importance of the research system to take up the challenges of the Sustainable Development Goals and offer incentives for doing so. Such a compelling statement would reverberate through the current evaluation system and, hopefully, contribute towards making research evaluation more compatible with developmental goals.

7. Concluding remarks

It is hard to imagine and even harder to implement alternatives to the prevailing global research evaluation system. Research evaluation is the outcome of complex social processes and involves powerful interest groups favoring the status quo. Designing feasible alternatives to such an entrenched system thus needs to consider power relations. It is difficult to assess if the proposal outlined in this article would have stakeholders strong enough to consider it seriously. A particularly difficult point is to acknowledge the differences between evaluating research outputs and evaluating 'the process of planning research that is aimed at social impact, as well as how this changes the power dynamic'.²

On the one hand it is not too confrontative, by allowing flexible adherence to either mode of research evaluation; on the other hand, by being flexible, it devaluates the exercise of power by academic elites and governments over the 'academic masses'. To address the challenges posed for the North and South by the SDG, public policy should recognize that it is high time to give precedence over measuring, controlling, punishing, selecting, comparing and ranking, to encouraging and facilitating the research-society dialogue. Some would say that this would lead to 'second rate science'; many more, including an increasing share of those working in public policy, will probably say that the outcome will be a better match between

research and the achievement of developmental goals. Working towards the compatibility of development and research evaluation is indeed a difficult challenge; it seems that many researchers, all over the world, are uniting and organizing themselves to face it. If this is so, public policy has a good chance to follow suit, giving opportunities for alternative research evaluation practices.

Notes

1. We thank an anonymous reviewer for highlighting the importance of taking all these aspects into account in research evaluation.
2. We thank an anonymous reviewer for this valuable remark.

References

- Altbach, P. (2003) 'Centers and Peripheries in the Academic Profession: The Special Challenges of Developing Countries'. In: P., Altbach (ed.) *The Decline of the Guru. The Academic Profession in the Third World*, pp. 1–21. Palgrave Macmillan: New York.
- Alvesson, M., and Spicer, A. (2016) '(Un)Conditional surrender? Why do professionals willingly comply with managerialism?', *Journal of Organizational Change Management*, 29/1: 29–45.
- ANII (2012) *Evaluación Del Impacto Del Sistema Nacional De Investigadores*, 6: 2018. <<http://www.anii.org.uy/institucional/documentos-de-interes/4/informes-de-evaluacion/>>, accessed January
- Arocena, R. (2015): 'La autonomía de la 'Universidad para el Desarrollo'', *Universidades, UDUAL, México*, 66: 7–18.
- , Göransson, B., and Sutz, J. (2018) *Developmental Universities in Inclusive Innovation Systems. Alternatives for Knowledge Democratization*. Palgrave Macmillan: New York.
- Bianco, M., Gras, N., and ——— (2016) 'Academic Evaluation: Universal Instrument? Tool for Development?', *Minerva*, 54/4: 399–421.
- Bölling, M., and Eriksson, Y. (2016) 'Collaboration with society: The future role of universities? Identifying challenges for evaluation', *Research Evaluation*, 25/2: 209–18.
- Borg, H., Galema, A., Mulder, H., and Steenbeek, S. (2016) 'The University of Groningen: an Engaging University'. In J., Goddard, H., Hazelkorn, L., Kempton and P., Vallance (eds.) *The Civic University. The Policy and Leadership Challenges*, pp. 201–221. Edward Elgar: London.
- Bozeman, B., and Sarewitz, D. (2011) 'Public Value Mapping and Science Policy Evaluation', *Minerva*, 49/1: 1–23.
- Brink, C., and Hogan, J. (2016) 'Newcastle University and the Development of the Concept of a World-Class Civic University'. In: J., Goddard, H.,

- Hazelkorn, L., Kempton and P., Vallance (eds.), *The Civic University. The Policy and Leadership Challenges*, pp. 240–256. Edward Elgar: London.
- Butler, L. (2007) 'Assessing University Research: a Plea for a Balanced Approach', *Science and Public Policy*, 34/8: 565–74.
- Council of Canadian Academies (2012) Informing Research Choices: Indicators and Judgment http://www.scienceadvice.ca/uploads/eng/assessments_and_publications_and_news_releases/science-performance/scienceperformance_fullreport_en_web.pdf, accessed 15 January 2018
- Dahler-Larsen, P. (2015) *Spazio Filosofico*, 1/13: 21–36. 'The Evaluation Society: Critique, Contestability and Skepticism', /ISSN: 2038-6788
- (2013) 'Evaluation as a Situational or a Universal Good? Why Evaluability Assessment for Evaluation Systems is a Good Idea, What it Might Look Like in Practice, and Why it is not Fashionable', *Scandinavian Journal of Public Administration*, 16/3: 29–46.
- DORA (2012) Declaration on Research Assessment, <https://sfedora.org/read/>, accessed 15 January 2018
- European Commission. (2012). *Responsible Research and Innovation. Europe's Ability to Respond to Societal Challenges*. https://ec.europa.eu/research/swafs/pdf/pub_public_engagement/responsible-research-andinnovation-leaflet_en.pdf, accessed 15 January 2018
- Evans, P. (1995) *Embedded Autonomy. States and Industrial Transformation*. Princeton University Press: Cambridge, MA.
- Foro Consultivo Científico Tecnológico y Academia Nacional Mexicana de Ciencias. (2005) Una Reflexión sobre el Sistema Nacional de Investigadores a 20 Años de su Creación. http://www.coniunctus.amc.edu.mx/libros/20_sni_final.pdf, accessed 15 January 2018
- Freeman, C., and Soete, L. (2009) 'Developing Science, Technology and Innovation Indicators: What We Can Learn from the Past', *Research Policy*, 38/4: 583–9.
- Gabriel, Y. (2005) 'Class Cages and Glass Palaces: Images of Organizations in Image-Conscious Times', *Organization*, 12/1: 9–27.
- Gläser, J., and Laudel, G. (2016) 'Governing Science', *European Journal of Sociology*, 57/1: 117–68.
- (2010) 'From Governance to Authority Relations?' In: R., Whitley, J., Gläser, and L., Engwall (eds.), *Reconfiguring Knowledge Production*, pp. 357–369 Oxford University Press: New York.
- Goddard, J. (2009) Reinventing the Civic University, NESTA, UK https://www.nesta.org.uk/sites/default/files/reinventing_the_civic_university.pdf, accessed 18 January 2018
- Halfman, W., and Radder, H. (2015) 'The Academic Manifesto: From an Occupied to a Public University', *Minerva*, 53/2: 165–87.
- Hammarfelt, B., and de Rijcke, S. (2015) 'Accountability in Context: Effects of Research Evaluation Systems on Publication Practices, Disciplinary Norms, and Individual Working Routines in the Faculty of Arts at Uppsala University', *Research Evaluation*, 24/1: 63–77.
- Hazelkorn, E. (2007) 'How do Rankings Impact on Higher Education', OECD Programme on Institutional Management in Higher Education. www.oecd.org/education/imhe/39802910.pdf, accessed 18 January 2018
- (2009) 'Impact of Global Rankings on Higher Education Research and the Production of Knowledge', *UNESCO Forum on Higher Education, Research and Knowledge, Occasional Paper*, 18:
- Hess, D. (2007) *Alternative Pathways in Science and Industry. Activism, Innovation, and the Environment in an Era of Globalization*. The MIT Press: Cambridge MA.
- Hicks, D. (2013) 'One Size Doesn't Fit All: On the Co-evolution of National Evaluation Systems and Social Science Publishing', *Confero*, 1/1: 67–90.
- (2012) 'Performance-based University Research Funding Systems', *Research Policy*, 41/2: 251–61.
- , Wouters, P., Waltman, L., et al. (2015) 'The Leiden Manifesto for Research Metrics', *Nature*, 520/7548: 429–31.
- Jonkers, K., and Zacharewicz, Y. (2016) *Research Performance Based Funding Systems: a Comparative Assessment*, Joint Research Center, European Commission. DOI: 10.2760/70120
- KNAW (The Royal Netherlands Academy of Arts and Sciences) (2011) *Quality Indicators for Research in the Humanities* <https://www.knaw.nl/en/news/publications/quality-indicators-for-research-in-the-humanities>, accessed 15 January 2018
- Kronman, U. (2013) 'Managing Your Assets in the Publication Economy', *Confero*, 1/1: 91–128.
- Magala, S. (2009) *The Management of Meaning in Organizations*. Palgrave Macmillan: New York.
- Martin, B., and Whitley, R. (2010). 'The UK Research Assessment Exercise: A Case of Regulatory Capture?' In: R., Whitley, J., Gläser, & L., Engwall (eds.), *Reconfiguring Knowledge Production*, pp. 51–79. Oxford University Press: New York
- Milanović, B. (2016). *Global Inequality: A New Approach for the Age of Globalization*. Harvard University Press: Cambridge, MA.
- Molas-Gallart, J., and Castro-Martínez, E. (2007) 'Ambiguity and Conflict in the Development of 'Third Mission' indicators', *Research Evaluation*, 16/4: 321–30.
- Neff, M. (2017) 'Publication Incentives Undermine the Utility of Science: Ecological Research in Mexico', *Science and Public Policy*, scx054, <https://doi.org/10.1093/scipol/scx054>
- OECD (2009) 'Enhancing Research Performance through Evaluation, Impact Assessment and Priority Setting', <http://www.oecd.org/sti/inno/Enhancing-Public-Research-Performance.pdf>, accessed 18 January 2018
- Ostrom, E. (1996) 'Crossing the Great Divide: Coproduction, Synergy, and Development', *World Development*, 24/6: 1073–87.
- PEDECIBA (2004) Criterios, herramientas y procedimientos generales para la evaluación de la actividad académica de los investigadores. Available at: <http://www.pedeciba.edu.uy/docspd/CritEvalInv04.pdf>, accessed 26 May 2018.
- Råfols, I., Molas-Gallart, J., Chavarró, D., and Robinson-García, N. (2016) 'On the Dominance of Quantitative Evaluation in 'Peripheral' Countries: Auditing Research with Technologies of Distance', paper presented at *Excellence Policies in Science*, Leiden 2–3 June. <https://ssrn.com/abstract=2818335>, accessed 15 January 2018
- Sarewitz, D. (2017) 'Kill the Myth of the Miracle Machine', *Nature*, 547/7662: 139.
- Sen, A. (1999) *Development as Freedom*. Oxford University Press: Oxford.
- Stiglitz, J. (2007) 'Prizes, Not Patents', *Post-Autistic Economics Review*, 42: 46–7. <http://www.paecon.net/PAERreview/issue42/Stiglitz42.htm>, accessed 18 January 2018
- Trencher, G., Yarime, M., McCormick, K., et al. (2013) 'Beyond the Third Mission: Exploring the Emerging University Function of Co-creation for Sustainability', *Science and Public Policy*, 41/2: 151–79.
- UN (2015) Sustainable Development Goals <http://www.un.org/sustainabledevelopment/sustainable-development-goals/>, accessed 1 January 2018
- Universities UK (2017) 'The Economic Impact of Foreign Students', <http://www.universitiesuk.ac.uk/policy-and-analysis/reports/Documents/2017/briefing-economic-impact-international-students.pdf>, accessed 15 January 2018
- Whitley, R. (2010). 'Reconfiguring the Public Sciences: The Impact of Governance Changes on Authority and Innovation in Public Science Systems'. In: R., Whitley, J., Gläser, & L., Engwall (eds.), *Reconfiguring Knowledge Production*, pp. 3–47 Oxford University Press: New York
- Williams, T., and Morrone, J. (2018) 'Science is Strengthened by Mexico's Researcher Evaluation System: Factual Errors and Misleading Claims by Neff', *Science and Public Policy*, doi: 10.1093/scipol/scy004.
- Wilsdon, J., Allen, L., Belfiore, E., et al. (2015) The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management, DOI: 10.13140/RG.2.1.4929.1363 http://www.hefce.ac.uk/media/HEFCE,2014/Content/Pubs/Independentresearch/2015/The_Metric_Tide/2015_metrictideS2.pdf, accessed 15 January 2018.
- Zaccai, E., Timmermans, B., Hudon, M., et al. (2016) L'évaluation de la recherche en question(s), Académie Royale des Sciences, des Lettres et des Beaux-Arts, Belgique.