

V. Concluding reflection and forward looking

The book provides a reflection on research assessment – what it is, why it is being carried out and who is performing it, in an attempt to understand how to best implement it. As it could be expected, we do not advocate one single way of conducting research evaluation and assessment. The excellence in the title should be regarded rather as *kaizen*, or continuous improvement of the way we comprehend and organise research endeavours, than a final blueprint.

We also argue that the excellent research assessment should contribute towards assisting the nexus of research stakeholders to reach a dynamic consensus on how to spend limited public resources towards the end of having plausible new scientific discoveries, solve pressing societal and business problems and create sustainable wealth.

We believe the current time calls for a reset and a reimagining of research evaluation activities, keeping the lessons learnt from the past in mind and looking forward to a new economic world. As of today, research evaluation is an immanent part of research policy and political priorities. However, we might want to see research evaluation as a complimentary self-reflection for the academia itself.

By shifting the focus from the (research) outcomes towards the process (of research discovery) we might be able to bring research evaluation to *kaizen*. Embedding innovative experimental participative forms of assessment, will certainly change the general picture of the assessment process and the system will become more flexible whilst losing some of its rigidity and bureaucracy.

A lot of researchers doubt whether the current governance of research lead to “higher quality” of research (Grande et al, 2013; Finke, 2014). Concerns are related even with the very “holy grail” of the current academic publishing – the “double blind review”. It often limits the academic dispute “behind closed doors” and influences the research production through the researcher expectation of publishability of results, rather than their genuine academic judgement. The very academic process is outcome-oriented, rather than process-oriented and hence the research evaluation systems tend to replicate that. Often even the conferences are judged rather on if their proceedings are indexed in Web of Science or Scopus and not on

what debate would happen there. Book projects usually provide greater flexibility, yet not all universities favour them for academic progress.

We conducted a comprehensive comparative analysis of research evaluation practices and their interdependence with the national innovation ecosystem in eight European Union countries. By doing so, we wanted to understand how the existing institutions frame research assessment systems and whether there is a difference in utilizing them as a political control mechanism (Shore & Wright, 1999), and whether this improves the output quality of the research system.

We employed a holistic approach in the study and a multi-vector (or multi-functional) system in order to present the research evaluation systems in the selected countries, and analysed their specific impact on the innovation ecosystem. To a certain extent the novelty in our approach allowed us to reach the conclusion that research evaluation is indeed used as a control mechanism, but not by the policymakers, but rather by the internal (to the universities) power networks. These networks are usually well-balanced and external (to the universities) reforms happen rarely. These reforms are most often internalised differently depending on which network is in power. This difference is higher in countries with lower law-enforcement.

We argued that research assessment should be considered within a managerial, economic, social, and environmental sustainability. A metaphor which helps us grasp the role of research assessment in the academic landscape is *gravity*. It dominates and governs research behaviour. It is conservative and preserves the status-quo in the long-run. Thus, a lot of progressive researchers are critical towards research assessment systems as they serve as additional gravity-centres and prevent quick changes sought by them – i.e., decoupling the academic career from the lengthy and expensive publishing process, democratizing the higher education process in a manner similar to the secondary education in Finland or, for example, in public schools in Maryland, USA. Progressive researchers would seek reforms, which will decrease inequality among universities and research centres originating from funding based on expensive publications.

Gravity, at the same time, helps organize chaos. In post-socialist societies the newly adopted research assessment systems, which mimic European models, served as a transformative power – dispose of the old gurus, who were gravitating around old, centralized models or at least change their orbit to have a western focus. At the same time, local systems were already

conservative, so they responded by negotiating with the major reference databases to include local outlets as well.

Countries like Austria and the Netherlands (or more generally those with traditions in market economy, democracy and civil society development) would tend to be more holistic and include criteria such as sustainable economic development, societal relevance, and viability in their research assessments (on every level: institutional, regional, or national). The holistic approach would require the involvement of civil society, and research performance would be evaluated also against the user's satisfaction (of research results). Holistic gravity tends to have strange attractors and less chaos and anomalies (for more details of chaos theory metaphors in management see Gilstrap, 2005).

By contrast, the new member states tend to focus on the abstract value of research measured by quantitative indicators related to publications only. Poland differs from the other CEEs by being closer to the holistic countries. In the book we discussed various potential explanations of those observations, among them the overall implementation of the rule of law or the lack of coherent industry-academia partnerships. Moreover, the distinction in the degree of holism could be attributed to the differences in participatory engagement of the overall policymaking. Last but not least, research assessment is both an instrument of control within the academic institution and it is also being shaped by the overall governance of academic institutions.

Countries and institutions which rely substantively on the whole Web of Science and Scopus and prioritize only quantitative indicators tend to produce more chaos, since gravity-centres emerge randomly and generate anomalies (i.e. researchers with relatively high scientometrics – close to 100 publications in Web of Science and Hirsch index=7, several doctoral degrees and professorships and at the same time – widely used plagiarism as a publication strategy and low overall quality of publications). Top universities in Germany and the UK for instance, maintain their own lists of recommended journals as publication outlets, quite shorter than the respective WoS/Scopus lists. The local conservative research community found easy strategies to publish a lot in the sacred databases without much impact or even negotiated cross-reference, which would be a proxy for impact.

Research evaluation as a holistic endeavour with a structured mix of different activities incorporates non-typical scientific activities which, however, are an immanent part of the research process itself. Such activities

include communication with different stakeholders, including societal, political, and the media. Thus, we did the same – we talked to policymakers, researchers, NGO activists, science communicators, and journalists. We found a great deal of dialectics in terms of which indicators signals what in different contexts.

Research evaluation could and should serve as a basis for a redesign of policies, if necessary, and increase research accountability to society. Last but not least, it provides legitimization for the resources invested. In this sense, its intervention in the innovation system could be quite tangible and could normatively contribute to its improvement. However, in certain cases the research evaluation system creates incentives for those involved to focus on its maintenance and reiteration, rather than on achieving *kaizen*.

So, in principle effective research evaluation and assessment should contribute to a higher level of coherence of research programmes, research performance, and societal impact. There is no clear high level of coherence in new member states due to the fact that the holistic approach in research evaluation is not used by them. One of the reasons for that could be fragmented and incomplete national innovation ecosystems consisting of smaller sub-ecosystems where stakeholders form isolated cliques instead of interlocking dense networks.

The German system of career development through various universities provide an excellent interlocking governance, but it is unclear how we could transform one system (shorter pre-career periods, inbreeding, high share of tenured lecturers, difficult lay-offs, low basic salary with virtually no performance based payment, commercialising reputation in business or projects) into another (longer and more diversified academic paths, high diversity in payment, high share of non-tenured lecturers and assistants, performance based requirements).

If we do want to change the way research evaluations work and impact the strategic planning of research organisations, we might want to involve the new generation of researchers who are still not burdened by traditions or historical overlays. Revolutionary changes always come from the young. They can change the rigidity of the system, make it future-oriented, and engage more and new stakeholders and channel their value propositions towards the research system. Attracting new stakeholders into the evaluation process will cease its momentous performance because the quality and impact of the research evaluation process depends not only on the main performers but also on all supporting staff involved in the process and on the audience, which should be adequately precluded in advance. We

could talk about evaluation curators from the early stage of preparation of the evaluation. The same way as curators redefine art and add value to the exhibition or performance, those research evaluation curators could be the masters of this collective priority-setting mechanism. Using the 12 art roles (Kolev, 2023) we can enrich the assessment stakeholder diversity by instrumenting the patron (companies), connoisseur (anti-plagiarism technologies) and others.

Activities related to research assessment have an impact on the development of the innovation ecosystem. In some cases, the intervention is effective and leads to a clear positive development trend. In other cases, the intervention is more sporadic, and this does not lead to the anticipated effect on the innovation ecosystem.

An important factor for the lack of coherence between research programs, research performance and societal impact in Eastern European countries is the strong insider influence on how the external research evaluation programs are internalized. While in Western academic landscapes we can partially accept the criticism of audit culture by Shore & Wright (1999), in Eastern academic landscapes and Bulgaria in particular, the research evaluation systems favour inertia.

In order to enhance the overall effect of the impact on the innovation ecosystem, the research assessment criteria system should be changed in two aspects: reviewing/bringing main indicators up-to-date and introducing adequate impacts for the individual criteria. Also, it is possible to create a criteria system which is similar to the ones in other countries where there are two types of indicators: compulsory and additional. Each of these groups deals with specific impacts relevant to the potential impact of a given indicator.

The research assessment has to be conducted based on the highest ethical standards and in good faith both on the part of the assessors and the entities providing the information for the assessment, in order to guarantee the usefulness of this specific activity for the units which are being assessed and the policymakers. Ethical standards should be inherently linked to a clear guarantee for lack of plagiarism.

For the purpose of improving the innovation ecosystem, it is necessary to monitor and control the assessment processes and their impact on quality by taking into consideration the interests of the stakeholders and by maintaining an open dialogue with them.

As we live in turbulent times, the research institutions at all levels – department, organisation, regional or national level – should be ready to

react to external shocks. First, we have had to face various crises – Covid-19, extreme floods, Russia-Ukraine war and so on. Despite the existence of various scenarios for all such events, the research systems must quickly adapt, study, and propose advice in a fundamentally shorter timeframe than before. Second, universities and research organisations should behave as they are teaching and preaching; they should be more environmentally-friendly, socially-responsible, and open to societal problems the same way as they are open to business problems. So, research evaluation systems should be able to capture the research impact along those criteria as well.

Based on the analysis, we could derive the following three key principles in research evaluation, which are prerequisites for quality and impactful research evaluation leading to research excellence:

- transparency: sharing the preparatory work, progress performance and results;
- collaboration: intertwining research and societal stakeholders and policy processes;
- trust and integrity: enhancing the academic reputation and maintaining integrity of research evaluation processes.

The Eastern European transition towards market economies used significantly the newly formed civil society organisations. They experienced different influence – predominately the Anglo-Saxon way of organisation of civil society, however Germany, Netherlands and Austria also shaped them. The German political foundations resorted to supporting civil society in doing socially important research, rather than the universities. So, politically the Eastern European civil society has enough foreign reputation (not to mention in US and UK) to be included as a valuable stakeholder in the next generation research evaluation. Ironically, often development projects of civil society organisations led to academic publications. Independent researchers contributed to citizen science of higher international impact (number of publications, academic citations and reference in policy documents) even compared to top schools of social science in some countries.

We would outline the following observations for the system:

The periodic research assessment does not have a clear effect on the innovation ecosystem. Some elements thereof are implemented but no tangible positive and sustainable trend can be established. The elements influenced by the implementation of research assessment can be found in:

- different levels of steady internationalisation of research and its presence in international research networks;
- ambitions to increase participation in research projects- national and international with a different scope. This is a clear sign that the project culture, capacity and competitiveness of the research and innovation potential is increasing;
- research training of PhD students which practically leads to an improvement in the quality of the staff.

Based on the analysis, we could project that it is highly advisable for other small countries on the Balkans (they are either accession countries or in pre-accession status) to introduce an independent external assessment of the research system for a given period of time. In this respect, the past practices – albeit sporadic – confirm the usefulness of this type of assessment.

Subject to discussion is, however, to what extent findings of this kind will be taken into account and will be followed up by corrective actions for the purpose of filling the gaps in the system. The practice in other countries is varied, but in those cases where the national structures have followed the recommendations and have undertaken the relevant actions, the effect on the innovation ecosystem is tangible. In those cases where the recommendations are ignored, there is no change in the existing status of the innovation ecosystem.

The literature review identified various studies of the future of research evaluation. It is extremely quantitative, dominated by smarter and more intelligent bibliometric infrastructure (Krüger & Petersohn, 2022) and new alternative metrics, including text mining and integrated open data on which research is based (Wilsdon et al, 2017). Media-driven rankings such as Handelsblatt BWL ranking and Times Higher Education will continue to emerge and contribute towards the reputation of the universities and their research. Yet, what is missing in the existing visions of future research evaluation is its perception as an epistemic process in the context of different epistemic regimes (Bösch, 2019), which is also a collective priority-setting mechanism.

We also offer a rather eccentric proposal – gamification of research evaluation with a variable term for periodic assessments or a continuous assessment and participation of all different stakeholders – academia, business, civil society, and policymakers. After all, the research evaluation is a game of reputation, and we should employ techniques that effectively govern

gamification. In recent years the monopoly of academic researchers over the “academic knowledge” has been dissolved by having a lot of civil society activists, business experts and even policymakers publishing through the “double blind review” process in reputable outlets. So, research evaluation should be extended towards the whole ecosystem producing *knowledge* and engaging all of those, who produce and consume knowledge. If we want a democratising science we need to have a democratised research evaluation system.

Institutions like the National Science Foundation, the Joint Research Center of the European Commission, Max Plank Society and the Chinese Academy of Sciences have already experimented with artificial intelligence tools for scientific assessments. Definitely AI would attract significant attention of research assessment scholars in the future!

References:

- Böschen, S. (2019). Processing issues in science policy: Emerging epistemic regimes. In *Handbook on science and public policy* (pp. 317–335). Edward Elgar Publishing.
- Gilstrap, D. L. (2005). Strange attractors and human interaction: Leading complex organisations through the use of metaphors.
- Grande, E., Jansen, D., Jarren, O., Rip, A., Schimank, U., & Weingart, P. (eds). (2013). *Neue Governance der Wissenschaft: Reorganisation-externe Anforderungen-Medialisierung*. transcript Verlag.
- Finke, P. (2014). Citizen science. *Das unterschätzte Wissen der Laien*. München: oekom.
- Kolev, I. (2023). Defining art as phenomenal being. *Arts* 12:100
- Krüger, A. K., & Petersohn, S. (2022). From Research Evaluation to Research Analytics. The digitization of academic performance measurement. *Valuation Studies*, 9(1), 11–46.
- Shore, C., & Wright, S. (1999). Audit culture and anthropology: Neo-liberalism in British higher education. *Journal of the royal anthropological institute*, 557–575.
- Wilsdon, J. R., Bar-Ilan, J., Frodeman, R., Lex, E., Peters, I., & Wouters, P. (2017). Next-generation metrics: Responsible metrics and evaluation for open science.