Next Generation Metrics for Scientific and Scholarly Research in Europe

LERU position paper
April 2024

LERU

PUSHING THE FRONTIERS OF INNOVATIVE RESEARCH

Table of contents

Acknowledgements

Authors: Jeroen Bosman (Utrecht University), Koenraad Debackere (KU Leuven), William P. Cawthorn (University of Edinburgh), Paola Galimberti (University of Milan), Mikael Graffner (Lund University), Leonhard Held (University of Zurich), Karlijn Hermans (Universiteit Leiden), Fiona Killard (formerly Trinity College Dublin and The University of Dublin, at present The National Institute for Bioprocessing Research and Training, Ireland), Ignasi Labastida (Universitat de Barcelona), Andrew Millar (University of Edinburgh), Mark Robinson (University of Zurich), Kristell Roser (Sorbonne University), Terhi Sangren (University of Helsinki), Michael Svendsen (University of Copenhagen), and Chair Paul Wouters (Universiteit Leiden)

With thanks to the LERU colleagues who contributed: Bernd Kortmann, Rens Vliegenthart, Immo Warntjes.

Consolidated Overview and Recommendations

What this report is about

The field of evaluating academic activities is vast, complex, and highly dynamic, as are the roles of any data and indicators used to support these evaluations. This report explores how universities can and should use currently available metrics and data to assess their research processes, in conjunction with qualitative expertise and information. While this report does not comprehensively cover responsible research evaluation, we recognise the need for next-generation metrics as a crucial aspect of responsible research evaluation. Using metrics out of context simply does not make sense, regardless of their ease of availability.

As if the challenge of evaluating academic work were not complex enough, universities are also confronted with the problem that they often lack the required resources, competences, and experiences to deal with each dimension in an expert manner. So, they improvise and make do with what they have. To make matters worse, many university administrative systems and processes are not able to easily produce validated, reliable, and meaningful data in-house due to fragmented data collection, incompatibility of datasets, lack of data standardisation, lack of maintenance, outdated ICT systems, and lack of expertise. As the LERU review of Open Science activities has shown, the simple act of monitoring what we are already doing requires a significant time investment and is mostly done manually.

In this context, we have chosen to focus this report on the aspect of academic evaluation that shows great potential for significant advancements in the coming years: the use and advancement of next-generation metrics for responsible research evaluation, encompassing open science, societal impact, and innovation. Addressing the crucial matter of next-generation metrics for academic teaching will necessitate the establishment of another LERU expert group.

We have chosen to create this report with the aim of maximising its utility for universities by compiling an extensive list of valuable resources on next-generation metrics. We also place strong emphasis on the imperative of collaboration, both among universities and between universities and funding agencies. By leveraging international expertise in bibliometrics and informatics, universities will enhance their resilience, reduce dependence on a limited number of companies that provide (seemingly) pre-packaged research intelligence, and gain a greater capacity to shape their metric policies in alignment with their own missions, rather than relying solely on standard metrics and data availability.

There is one overarching recommendation to universities and the League of European Research Universities (LERU):

Use indicators and metrics that are contextually relevant, that support responsible research evaluation, and that align with your institution's mission. Institutions should collaborate and reuse existing metrics expertise in order to maximise their efficiency in achieving this goal.

Defining next generation metrics

We propose that next-generation metrics may encompass both existing and new indicators, spanning all 'levels' and types, provided they align with the institutional goals and ambitions, demonstrate validity and robustness in accordance with the latest scientific insights, and are responsibly applied in accordance with the recommendations of frameworks developed over the past decade (refer to **Appendix I** for an overview of metrics terminology).

This definition includes the use of already existing and validated metrics in novel ways and/or within different contexts. Importantly, it also incorporates the insights gained from the experiences of previous decades in using metrics. Consequently, this report also delves into discussions on how *not* to use indicators, covering both their misuse and abuse. Depending on the level of general expertise within the academic community at the university, this may be the most crucial initial conversation to have. For instance, in those faculties or departments that still heavily rely on a single indicator for all evaluation purposes (a '*one Ring to rule them all*' scenario, such as the h-index).

We have identified two primary drivers that influence the demand for next-generation metrics:

- The necessity to 'measure things differently' stems from dissatisfaction with the use of current indicators and metrics in inappropriate contexts.
- The requirement to 'measure different things' arises from the new expectations placed on universities, such as open science, impact, societal relevance, and integrity.

Both aspects need to be considered in tandem, as they interact in shaping new research evaluation practices and the development of novel metrics.

Topical overview of metrics and policies

Our report focuses on four main areas:

1. An overview of the development of next-generation metrics, their use, and their limitations.
2. An overview of the current status of metrics policies at LERU universities.
3. A dynamic visualisation of the opportunities to leverage university data that connects existing and potential next-generation metrics.
4. Recommendations concerning next-generation metrics policies and evaluation practices.

To maximise the usefulness of this report, we present relevant recommendations at the conclusion of each chapter.

In terms of potential scenarios for LERU members, we reconfigure these recommendations and present them in this introductory chapter.

The first focus is covered in the initial eight chapters. The report commences by outlining the evolution of research metrics in the context of the interplay between academia and science policy. We delve into the significance of metrics, their essential nature, and their constraints, while also examining the origins of metric misuse. Subsequently, we explore the hurdles associated with a new generation of scientometrics, often referred to as Scientometrics 2.0. Next, we draw a comparison between the limitations of evaluative peer review and those of summative metrics, demonstrating how they can serve as complementary approaches.

One of the challenges in using metrics is their frequent application at the incorrect level of aggregation and analysis within the university. They often 'trickle down' from higher policy levels to the scientists' workbenches. To address this issue, we focus on 'phenomenon-specific' considerations in a dedicated chapter. This is followed by an examination of a closely related and current topic in our discussions: the utilisation of next-generation metrics and university rankings in assessing and communicating academic research performance.

We conclude our examination of next-generation metrics with a discussion of the ethical and political dimensions, including how to address the emerging ethical concerns (both opportunities and challenges) related to artificial intelligence-based metrics.

Chapter 9 comprises the second focal point of our report: What are LERU universities currently implementing with regard to next-generation metrics, and what is the level of our awareness concerning all relevant issues related to them? In other words, how well-versed are we in metrics? Given the time constraints of our work in this expert group, it was not feasible to encompass all aspects of metrics use and policies in this overview. Therefore, we opted to concentrate on the dimension that is currently generating the most enthusiasm within our universities: the development of open science policies. We acknowledge that this may not provide a comprehensive view of metrics literacy across all application domains, but it does offer an insightful snapshot of one of the most dynamic dimensions of our practices and policies.

This overview is derived from a comprehensive examination of the university documentation available to us and from a survey we conducted among all LERU members. Additionally, we incorporated insights from the review of the LERU Roadmap for the implementation of Open Science in the LERU institutions, conducted by the LERU Ad Hoc Group Open Science Ambassadors. It is important to acknowledge the limitation of this approach.

The field of next-generation metrics is dynamic, and it is possible that we may have missed some important developments at individual universities. Nevertheless, when considered collectively, we have gained a reasonably robust understanding of the field's current status.

The survey results reveal a diverse array of activities and support associated with open science within LERU universities and the countries in which they are situated. It is clear that different universities and countries are at different stages of open science development, with some having more established systems for supporting and overseeing open science. This diversity may be attributed to variations in the particular needs each university or country seeks to fulfil and their perspectives on how open science can meet these needs. Moreover, the discrepancies can be influenced by unequal access to the requisite resources for supporting and monitoring open science.

At institutional, national, and international levels, it is evident that the most well-established open science policies and monitoring primarily pertain to Open Access to publications. This includes the maintenance of regularly updated dashboards for reporting the percentages of open-access publications. While some universities and countries have implemented policies related to FAIR Data, these are less widespread compared to Open Access. Regarding the other six pillars of open science (Education and Skills, Rewards and Incentives, Next-generation metrics, Research Integrity, Citizen Science, and The European Open Science Cloud), formal policies and monitoring have not yet been put in place. It is important to note that many of the eight open science pillars have interconnected and overlapping elements. For example, progress in Open Access and FAIR data can contribute to enhancing Research Integrity. Therefore, it might not be necessary to create separate policies for each pillar but, instead, consider how open science policies can collectively advance multiple pillars simultaneously.

The third focus of our report addresses the relatively new challenge of data availability for research evaluation. As institutions introduce new values, goals, and policies, there is a growing demand for data to inform, assess, and monitor their progress. Given the diversity of values, goals, and policies across research disciplines, teams, departments, institutions, and countries, it becomes clear that a one-size-fits-all, non-customisable metric cannot adequately meet the current demand. Simultaneously, we are witnessing an increase in aggregators offering machine-readable knowledge graphs with granular data, aiming to comprehensively capture research activities and outputs. These services and tools provide large amounts of data, but they exhibit substantial variations in data quality and even more diverse degrees of completeness in their records compared to traditional, narrower, and more selective databases. Furthermore, they necessitate more comprehensive documentation to make sense of the data and address biases and data gaps.

How can universities best address this challenge? We suggest that it is important to distinguish between:

1. Granular data (mainly comprising characteristics of researchers, organisations, and their research activities and outputs).
2. Metrics (involving calculations and combinations of data for specific use cases, allowing comparisons over time, against a baseline, or among individuals, organisations, or other groups, summarised into validated and robust indicators).
3. Algorithms, code, and software used to generate or analyse data.

In this report, we emphasise the granularity of data because we believe universities can make the most significant progress in this area. After all, it is their data. To complement the recommendations and inspire universities, we have used a dynamic visualisation of the potential links between data and metrics on the one hand and sought-for changes in measurement on the other, which is summarised in **Tables 2** and **3** and **Figure 3**, with full detailed versions in Zenodo[1].

Currently, only a limited number of universities can fully leverage their data in these ways. However, this situation is not expected to persist. Therefore, we propose that universities could benefit from collaborating to facilitate advancements in this domain. Even if universities choose to continue outsourcing their research intelligence capabilities, such collaboration would enhance their in-house expertise, enabling them to better assess which commercial services to use (and which to avoid or discontinue).

Perhaps a more fundamental reason for this approach is that open science practices also require transparent policies regarding metadata and data infrastructures, all while respecting legal and privacy constraints.

You are not alone: use the available resources!

Throughout the report, we have referenced an extensive body of pertinent literature and readily available resources. The positive news is that we are witnessing a rapid increase in the number of universities and funding agencies making the shift from using overly simplistic indicator-based research evaluation and monitoring methods (including the anxiety of unexplained declines in rankings) to adopting a framework for responsible research evaluation. This transformation has given rise to numerous innovative initiatives and experiments to aid the transition to the development and use of next-generation metrics.

1 Jeroen Bosman. (2024) Changes in demand and supply of metrics for research evaluation in the context of open science and new R&R. https://doi.org/10.5281/zenodo.10569960

In this report, we emphasise three distinct categories of resources that are accessible to universities:

1. Experiments at universities involving innovative approaches in responsible research evaluation, next-generation metrics, and open science initiatives. We highly recommend expediting this process by sharing knowledge and being willing to take calculated risks.
2. In-house metrics expertise within universities (e.g., at Leuven, Leiden, UCL, DZHW, Charité Berlin).
3. Policy frameworks and research literature on the shift to responsible research evaluation.

In this context, we recommend that LERU consider the INORMS SCOPE Framework for Research Evaluation[2] as their comprehensive guide (which does not imply that every university must adhere to the framework in every detail).

The fourth focus revolves around the numerous recommendations presented throughout the report, which are summarised below.

Key recommendations

The 36 key recommendations this study produced provide guidance in the transition towards a more contextually relevant and diverse set of metrics and indicators to support responsible summative and formative evaluations that align with institutional goals and missions. Some are aimed at university leadership and research directors, some more at specialists involved in research evaluation and research intelligence and some primarily at the research community at large. We refrained from grouping recommendations by these and other ‘target groups’ because we feel that it is up to institutions themselves to allocate responsibilities and activities, as well as to prioritise recommendations to engage with.

Recommendations for the transition to next generation metrics

(For context see chapter 1)

1. To initiate the transition toward the adoption of next-generation metrics in LERU institutions, the first crucial steps involve conducting a comprehensive assessment of the existing metrics, identifying their strengths, weaknesses, and areas for improvement.
2. Involve all relevant stakeholders within your community in the decision-making process. Evaluate with the evaluated.
3. Metrics should extend beyond traditional bibliometrics and encompass societal impact, open science, collaboration, and recognition of diverse contributions to research and academia.
4. Establish clear objectives and goals for the transition. These goals should closely align with the institution's mission, values, and strategic priorities to ensure that the adoption of next-generation metrics is in harmony with the institution's overall vision.
5. The institution can then develop and implement pilot projects to assess the use of next-generation metrics in specific departments or research groups.
6. Provide training and support to Faculty and researchers to ensure they understand how to use and interpret these metrics effectively.
7. Over time, continuously evaluate the adopted next-generation metrics to ensure their ongoing effectiveness and relevance.

Recommendations on metrics policy

(For context see section 2.1)

8. Universities should explicitly consider the multilevel nature of both policies and metrics when developing and using next-generation metrics.
9. To ensure transparency and mitigate potential biases, full openness of data and indicators used in quantitative methods and metrics-based policies is crucial. This transparency allows stakeholders to gain a clear understanding of the metrics' limitations and potential for misuse.
10. The design and implementation of funding formulas and policy instruments should involve the active participation of all stakeholders, including the scientific community. This inclusive approach ensures a bottom-up perspective, fosters trust, and encourages open dialogue.
11. The SCOPE framework for research evaluation, developed by the international network of research management societies (INORMs), offers a valuable framework for universities, based on the principle of ‘evaluate with the evaluated’.
12. While using bibliometric indicators for policy purposes, stakeholders must remain aware of potential distortions that may arise.
13. Critically evaluate the biases inherent in the data and indicators used for policy decisions. This assessment will help in appropriately using bibliometric data and making informed policy choices.
14. Develop and adhere to ethical guidelines for the use of scientometric data to prevent misuse and safeguard the interests of researchers and institutions.

Recommendations on metrics use

(For context see sections 2.2, 2.3 and 5.1)

15. Utilise the technical expertise available in international centres of excellence, such as the Research on Research Institute (RORI), scientometric centres in Europe, and university libraries.

2 SCOPE Framework for Research Evaluation. https://inorms.net/scope-framework-for-research-evaluation/

16. Metrics should acknowledge and reward the contributions of teams, not solely individuals.
17. When developing and implementing next-generation metrics, institutions must be highly attentive to the contexts in which these metrics will be extracted and applied. Metrics should be designed and adapted to suit the specific levels of the academic system, taking into account the diversity of interactions and complexities involved.
18. Employ mathematical techniques like normalisation and standardisation to address biases and ensure fair comparisons in bibliometric data.
19. Promote open communication between the scientific community and policymakers to address concerns, promote understanding, and prevent potential misuse of scientometric data.
20. Recognise the complementary nature of peer review and metrics in assessing research impact and knowledge creation.
21. Understanding subject-specific communication behaviour within various publication types and channels is essential.
22. Avoid linking funding and allocation to performance metrics at higher levels of aggregation, and instead, link them solely to the mission and goals at the same level. This approach can help prevent the inappropriate trickling down of metrics and avoid negative consequences.

Recommendations on data handling

(For context see section 3.1)

23. Develop and adhere to ethical guidelines for the use of scientometric data to prevent misuse and safeguard the interests of both researchers and institutions.
24. Collaborate with researchers and data analysts to develop innovative methods for addressing big-data-related issues that arise from the integration of new data sources.
25. To address meso- and micro-level issues such as co-authorship, gender, and open-access (OA) publication, data processing approaches should focus on higher granularity, providing more detailed information for accurate analysis.
26. Enhancing and standardising elements of university registration systems can ensure the availability of relevant data for metric generation while ensuring the reliability and consistency of the generated metrics.

Recommendations on university rankings

(For context see chapter 7)

27. Develop guidelines for interpreting rankings, including approaches that consider emerging criteria such as sustainability and open science.
28. Institutions should be transparent about their use of rankings, explaining the purpose and intentions behind their application.
29. Policymakers and institutions should use the INORMS/SCOPE framework for evaluation to assess ranking systems and ensure they align with their goals and priorities.

Recommendations on ethics

(For context see section 8.1)

30. When developing new metrics, it is essential to prioritise ethical considerations. Metrics should be designed and used in ways that respect data privacy, intellectual property, and research integrity.
31. While next-generation metrics can be valuable tools in evaluating research performance and impact, they should not be the sole basis for evaluation. Universities should adopt a holistic approach that considers other factors, such as the quality and originality of research, real-world impact, and an academic's contributions to their field.
32. The recent emergence of artificial intelligence tools creates a whole new set of ethical and technical challenges and dilemmas. A proactive effort to explore this new territory as soon as possible would be prudent.

Recommendations on communication about open science and metrics

(For context see section 9.4)

33. All LERU members should maintain websites that detail institutional and national policies for open science, encompassing Open Access (OA) and FAIR data. These websites should be available in multiple languages, including English, to ensure accessibility.
34. All countries should establish open science dashboards similar to those from institutions like Charité Berlin or the French Open Science Monitor. These dashboards should report on various aspects, including OA publications, preprints, open data, preregistration, and more.

Recommendations on collaboration within LERU

(For context see chapter 1)

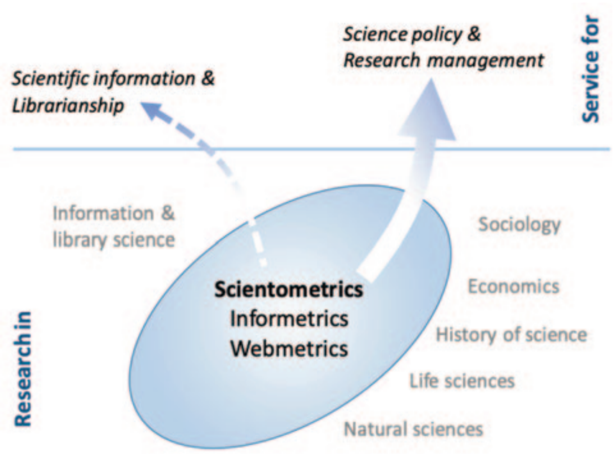
35. LERU members should collaborate to exchange best practices and code for handling and analysing open data relevant for research evaluations, and provide open access to diverse and pertinent data sources, making it easier to calculate flexible next-generation metrics.
36. Foster partnerships with funding organisations, for instance, through the Research on Research Institute and by connecting with the European Coalition on Reforming Research Assessment (CoARA), which is dedicated to achieving a more inclusive and responsible research assessment.

1. Scientific communication and the transition to next generation metrics

Science policy leadership and academic leadership have long been interested in understanding and assessing scientific and scholarly activities (Figure 1). The inputs and outputs of these activities are essentially knowledge coded as data and information, making the measurement of scientific communication and derived metrics a priority on policy and leadership agendas. The scientific literature plays a crucial role in consolidating knowledge and information from scientific endeavours. The rise of literature-based metrics, however, has been marked by both useful applications and potential as well as actual misuses. In the current context, the pursuit of novel metrics should be guided by an awareness of the lessons learned from these historical assessments of scientific output and impact.

Publications in journals are a critical communication channel in the fields of biomedical, life and natural sciences, and increasingly social sciences and humanities. Peer review, where experts use qualitative judgements to endorse the content, has become a fundamental mechanism for ensuring the integrity of these outputs and essential to scientific progress.

Figure 1. The symbiosis between bibliometrics and research management and the development of science policy.



Both the publication literature, which is often used in scientometric or bibliometric applications, as well as the peer review mechanism, have become keystones in the fields of science, technology, and innovation policy.[3]

Since the first half of the 20th century, journal papers have displayed a high degree of predictability and standardisation in their bibliographic structure, rendering them an optimal subject for (semi) automated metadata extraction prior to statistical analyses. If size were the only concern, bibliometrics would be reduced to a mere counting game.

Garfield's SCI citation network in the 1960s provided yet another step forward[4]. By analysing citations in papers' bibliographies, the network enables the accurate measurement of the impact of documents/authors on the scientific community. It complements peer review, which can be biased, by ranking authors and institutions based on publication outputs and citation scores and mapping bibliographic links between scientific documents to aid in scientific merit allocation and understanding scientific communication.

Multi-level science policy has become an integral part of government and university functions since the publication of Vannevar Bush's influential work "*Science, the Endless Frontier*" in 1945. Along with the development of science policy, the discipline of scientometrics emerged to assist the scientific community in accessing, retrieving, and disseminating their ever-growing output. A multilevel portfolio of metrics and methodologies has developed, serving multiple target groups. This diversity of outputs and applications necessitates a proper understanding of the "dos and don'ts" of scientometrics.

To initiate the transition toward the introduction of next-generation metrics in LERU institutions, the first crucial steps involve conducting a comprehensive assessment of the current metrics. This assessment should identify the strengths, weaknesses, and areas for improvement, including any unmet needs within the existing metrics, thus providing a clear starting point for the transition. To ensure a successful adoption, various stakeholders including faculty, researchers, administrators, policymakers, and students, should be included in the decision-making process.

3 Debackere, K., Glänzel, W., Thijs, B. (2019). Scientometrics Shaping Science Policy and vice versa, the ECOOM Case. In: Glänzel, W., Moed, H.F., Schmoch, U., Thelwall, M. (eds) Springer Handbook of Science and Technology Indicators. Springer Handbooks. Springer, Cham. https://doi.org/10.1007/978-3-030-02511-3_17

4 P. Wouters (2006). "Aux origines de la scientométrie: La naissance du Science Citation Index", Actes de la recherche en sciences sociales 164 (4), 11-22.

A collaborative approach of this nature ensures a thorough exploration of the potential applicability of emerging next-generation metrics across various academic disciplines and research areas. It also allows for the tailoring of these metrics to address the specific needs and nuances of each academic domain. Metrics should go beyond traditional bibliometrics and consider societal impact, open science, collaboration and acknowledge diverse contributions to research and academia. In addition, the establishment of clear objectives and goals for the transition is necessary. These goals should be closely aligned with the institution's mission, values, and strategic priorities to ensure that the adoption of next-generation metrics is in harmony with the overall vision of the institution.

Fortunately, LERU universities do not have to create this anew. They can connect with the global movement to reform research evaluation[5] and to the European coalitions that are driving these reforms[6]. The principal themes of this development are “responsible research evaluation” and “responsible metrics”. This has led to the creation of resources, case studies, and exemplary initiatives, along with platforms for exchanging experiences. LERU universities are encouraged to take advantage of these resources.

After the collaborative consultation phase, the institution can develop and implement pilot projects to test the use of next-generation metrics in specific departments or research groups. During this phase, faculty and researchers can receive training and support to ensure their effective use and interpretation of these metrics. This also serves to raise awareness about the benefits and limitations of next-generation metrics. As a group, the LERU members can collaborate to share best practice and provide open access to diverse and relevant data sources, thus simplifying the calculation of next-generation metrics. This comprehensive approach will facilitate the successful integration of these metrics into the institution's research assessment framework, empowering the academic community to harness the full potential of next-generation metrics in a meaningful and informed manner.

Over time, the adopted next-generation metrics should undergo continuous evaluation to ensure that they maintain their effectiveness and relevance. They can be refined and adapted based on feedback and new developments. Creating strategies and feedback to the community that emphasises the benefits of these metrics in aligning research with societal needs and showcasing how they complement existing evaluation practices can help alleviate pressure from the scientific community and facilitate a seamless and successful transition.

5 Partnership (IAP), the InterAcademy. “‘The Future of Research Evaluation: A Synthesis of Current Debates and Developments,’” May 25, 2023. https://www.interacademies.org/news/future-research-evaluation-synthesis-current-debates-and-developments
6 CoARA. 2022. Agreement on Reforming Research Assessment. https://coara.eu/app/uploads/2022/09/2022_07_19_rra_agreement_final.pdf

2. The role of metrics in informing and supporting academic policy development

To inform and support academic policy, it is important to have high-quality scientometric data and indicators. This requires advanced research into developing a state-of-the-art indicator base, as well as an IT-system that is easily accessible, reliable, and able to support scientometric research and service tasks effectively and efficiently.[7] The interplay between data, indicators, and code is pivotal in shaping next-generation metrics. Access to diverse and relevant data sources, combined with the right indicators, ensures accurate and comprehensive measurements. Additionally, the code used in the calculation process plays a critical role in transforming raw data into meaningful metrics.

Scientometrics also plays an important ex-post role in various activities, including mapping cognitive structures, actor connectivity, and institutional performance in academic systems. It is also used to assess the scientific performance of institutional actors in national and regional science systems, as well as monitor multi-annual strategic plans and accompanying funding schemes based on the productivity and visibility of their most-recent research activities.

This interplay between ex-ante and ex-post roles creates an intense symbiosis between scientometrics and the development of science policy trajectories and instruments, allowing for the evolution, optimisation, and revitalisation of these instruments. As science policy instruments continue to evolve, the field of scientometrics is required to produce new, more relevant, valid, robust, and unbiased methods and indicators. This need is further reinforced by the increasing use of performance-based funding models for research organisations[8, 9].

2.1 Recommendations for responsible use of quantitative methods and metrics-based policies

Bibliometrics is shaping science policy, and science policy is shaping bibliometrics, with the potential for ever new metrics developments.

The use of quantitative methods and metrics-based funding models and policy instruments will always be subject to debate. It is important therefore, for academic and science policy makers and scientometricians to understand their limitations and potential for abuse.

- Universities should explicitly consider the multilevel nature of both policies and metrics when developing and using next-generation metrics.
- To ensure transparency and mitigate potential biases, full openness of data and indicators used in quantitative methods and metrics-based policies is crucial. This transparency allows stakeholders to gain a clear understanding of the metrics' limitations and potential for misuse.
- The design and implementation of funding formulas and policy instruments should involve the active participation of all stakeholders, including the scientific community. This inclusive approach ensures a bottom-up perspective, fosters trust, and encourages open dialogue.
- The SCOPE framework for research evaluation, developed by the international network of research management societies (INORMs), offers a valuable framework for universities, based on the principle of 'evaluate with the evaluated'[10].
- While using bibliometric indicators for policy purposes, it is essential for stakeholders to remain aware of potential distortions that may arise. These distortions can impact the publication, citation, and collaboration behaviour of scientists. Understanding the cognitive and social influences on problem-choice behaviour can help assess how bibliometrics for policy purposes may affect scientists' research choices. The use of both existing and next generation metrics thus requires a profound awareness and understanding of their potential as well as their limitations. It is important that both researchers and policy makers have a good insight into the benefits and shortcomings of the bibliometric instruments at their disposal.
- Studies on the problem-choice behaviour of academic scientists have revealed that both cognitive and social influences determine the way scientists go about choosing the problems they work on[11].

7 Debackere, K., Glänzel, W., Thijs, B. (2019). Scientometrics Shaping Science Policy and vice versa, the ECOOM Case. In: Glänzel, W., Moed, H.F., Schmoch, U., Thelwall, M. (eds) Springer Handbook of Science and Technology Indicators. Springer Handbooks. Springer, Cham. https://doi.org/10.1007/978-3-030-02511-3_17

8 Jonkers K. and Zacharewicz T. Research Performance Based Funding Systems: a Comparative Assessment . EUR 27837. Luxembourg (Luxembourg): Publications Office of the European Union; 2016. JRC101043. https://publications.jrc.ec.europa.eu/repository/bitstream/JRC101043/kj1a27837enn.pdf

9 Pruvot, Enora, Anna-Lena Claeys-Kulik, and Thomas Estermann. "Strategies for Efficient Funding of Universities in Europe," 153–68, 2015. https://doi.org/10.1007/978-3-319-20877-0_11

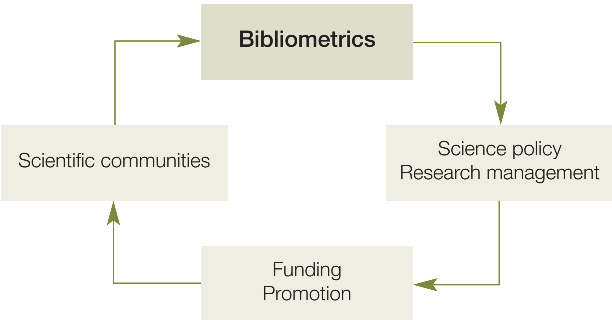
10 INORMS. SCOPE Framework for Research Evaluation. https://inorms.net/scope-framework-for-research-evaluation/

11 Debackere, Koenraad, and Michael A. Rappa. "Institutional Variations in Problem Choice and Persistence among Scientists in an Emerging Field." *Research Policy* 23, no. 4 (1994): 425–41. https://doi.org/10.1016/0048-7333(94)90006-X

It is necessary, therefore, to critically evaluate the biases inherent in the data and indicators used for policy decisions. This evaluation will help in using bibliometric data appropriately and making informed policy choices.

By implementing these recommendations, academic and science policy makers, as well as scientometricians, can work towards ensuring a more-informed, transparent, and balanced approach to the use of quantitative methods and metrics-based policies in the research and academic domain.

Figure 2. Scientometrics Shaping Science Policy and Vice-Versa



2.2 Addressing the improper use of scientometrics or bibliometrics

Improper use of scientometrics or bibliometrics can range from the unintentional to the deliberate, including selecting only the most favourable indicators or outright misuse of data. Bibliometricians themselves may also be responsible for uninformed use, as they may succumb to following popular trends in order to meet customer expectations. Any form of uninformed or deliberate misuse of bibliometric results could potentially harm the credibility of bibliometric research[12]. Uninformed use of bibliometrics can be seen in cases where indicators are incorrectly presented or interpreted, or when they are used in an inappropriate context due to insufficient knowledge of methodology, background, or data sources. Generalisation or induction of results obtained at lower levels of aggregation can also be a problem. Misuse of bibliometrics can take the form of intentionally incorrect presentation or interpretation of indicators, as well as their deliberate use in inappropriate contexts. Tendentious application of biases and selection of incompatible indicators are also examples of misuse.

The use of bibliometric tools can have unintended consequences even when used correctly. For example, the re-interpretation of citation practices may lead to authors avoiding self-citations, which could interfere with scientific communication. Furthermore, when bibliometric tools are used in decision-making for academic policy and research management, the scientific community may respond to feedback regarding their funding. The relevance and limitations of citation-based measures (including JIF, h-index etc.) should be standard knowledge in the academic and policy communities, however, this is not always the case, and this has been a point of criticism for performance-based funding formulas. It is important to regularly review and adjust the funding formula as new insights emerge regarding both its positive and negative effects. Butler (2004) demonstrated this in Australia, where linking funding to publication counts led to increased publication activity in lower-impact journals[13]. It is worth mentioning here that in Australia, journal rankings were created and used for evaluation purposes, but were later retracted due to being deemed ill-conceived. This system was criticised for pushing researchers to publish in certain venues, rather than those that were most appropriate for their research. This highlights the potential negative effects of the policy use of bibliometrics on the scientific community. As Figure 2 illustrates, there can be feedback loops between the two in both directions, which may have positive as well as negative impacts. Potential positive effects include encouraging scientists to recognise the benefits of scientific collaboration and publishing in top-tier journals (i.e. often perceived as those journals with a high journal impact factor) and stimulating their publication activity. However, there are also negative effects that need to be considered, including exaggerated collaboration and trends towards hyper-authorship, inflating publication output by dividing it into sequences (also known as "salami slicing"), inflating citation impact through self-citations and forming citation cliques, and prioritising visibility over quality and recognition[14].

12 Gingras, Yves. *Bibliometrics and Research Evaluation: Uses and Abuses*. Cambridge, Massachusetts: The MIT Press, 2016. https://doi.org/10.7551/mitpress/10719.001.0001
13 Butler, Linda. "Modifying Publication Practices in Response to Funding Formulas." *Research Evaluation* 12, no. 1 (April 1, 2003): 39–46. https://doi.org/10.3152/147154403781776780
14 Glänzel, Wolfgang, Pei-Shan Chi, and Koenraad Debackere. "Measuring the Impact of Research – from Scholarly Communication to Broader Impact." In: Rafael Ball (Ed.), *Handbook Bibliometrics*. De Gruyter Saur, 2020. ISBN: 978-3-11-064227-8. https://doi.org/10.1515/9783110646610-015

2.3 Recommendations to address improper use of scientometrics or bibliometrics

Fortunately, the increasing sophistication of bibliometric research has provided us with the necessary insights and tools to address the various positive and negative effects and instances of use and abuse. The following measures can be used to address these issues:

- Implement mathematical techniques like normalisation and standardisation to address biases and ensure fair comparisons in bibliometric data.
- Encourage open communication between the scientific community and policymakers to address concerns, promote understanding, and prevent potential misuse of scientometric data.
- Recognise the complementary nature of peer review and metrics in assessing research impact and knowledge creation. Utilise both of these tools effectively for monitoring and enhancing the scholarly process[15].
- Establish mechanisms for continuous monitoring and evaluation of the impact of scientometrics on academic and science policy, identifying any unintended consequences and adjusting strategies as needed.
- Develop and adhere to ethical guidelines for the use of scientometric data to prevent misuse and protect the interests of researchers and institutions.
- Promote interdisciplinary collaboration among sciento-metricians, researchers, policymakers, and stakeholders to create robust methodologies and indicators that encompass diverse perspectives.
- Raise awareness and provide education about the limitations and potential biases of scientometric data among researchers, administrators, and policymakers to ensure responsible and informed decision-making.
- Stress the importance of continually improving and adapting scientometric practices to keep pace with evolving research landscapes and societal demands.
- Leverage the technical expertise available in international centres of excellence, such as the Research on Research Institute (RoRI)[16], the scientometric centres in Europe, and university libraries.

By implementing these recommendations, the potential of scientometrics and bibliometrics can be harnessed while safeguarding against their improper use and ensuring a more comprehensive and responsible approach to knowledge creation and policymaking.

15 Glanzel,W., 2008. Seven myths in bibliometrics. About facts and fiction in quantitative science studies. In Kretschmer & F. Havemann (Eds.), Proceedings of WIS Fourth international conference on webometrics, informetrics and scientometrics and ninth COLLNET meeting, Berlin.
16 Research on Research Institute. https://researchonresearch.org/

3. The challenges for scientometrics 2.0

In 1969, the terms 'bibliometrics,' 'scientometrics,' and 'informatics' were coined to describe the application of mathematical methods to study communication and knowledge development. These concepts share common objectives, supporting information retrieval, structural analysis of knowledge, and informing research assessment and policy. Scientometrics was initially developed to model and measure the impact of research on scientific communities, focusing mainly on journal literature in basic sciences. It quickly became a tool for evaluating research performance, however, even though it was not intended to replace qualitative methods with quantitative methods. As evaluation processes became more dynamic, the focus of scientometric applications shifted from macro-level studies to meso-level and micro-level studies. This necessitated greater precision in data retrieval, but during the 1990s, the appropriate data sources and new methods were not yet fully available, leading to a sense of crisis in the field[17].

As scientometrics expanded its scope to incorporate new data sources, including conference proceedings and books, and broadened its focus to encompass research performance evaluation in diverse fields beyond basic research, it became evident that the original framework, centred on publication and citation, was too limited for these broader contexts. In response, scientometricians embarked on efforts to meet this challenge by expanding their scope and enhancing their methodologies to align with the evolving landscape of research evaluation.

The expansion of data sources and the broadening of scientometrics' scope resulted in two key characteristics. First, there was a 'perspective shift' as scientometricians adapted to new developments in research evaluation by broadening their scope and refining their methodologies. Second, there was a trend toward applications at lower levels of aggregation, moving from macro-level to meso-level analysis and increasingly toward the evaluation of individual scientists. This shift presented challenges in individual-level bibliometrics, as discussed by Wouters et al. (2013)[18]. Consequently, changes were observed not only in the diversification of targeted samples but also in the scale and scope of scientometric analyses.

3.1 Recommendations to address the challenges in scientometrics

The inclusion of new data sources and advanced features in scientometrics has brought forth several challenges that require novel solutions. The integration of new data sources such as proceedings, books, national databases, and web sources with traditional bibliometric data has led to big-data related issues, including name disambiguation, data cleaning, and the management of redundancies. Broadening the scope of scientometrics has also resulted in conceptual and methodological challenges that are closely tied to specific cultures in scholarly communication across various fields, such as the technology-related sciences, social sciences, and humanities.

To address these:

- Researchers and data analysts should work together to develop innovative methods to handle big-data related issues arising from the integration of new data sources. This may involve improving name disambiguation techniques, data cleaning algorithms, and redundancy management.
- Collaborative efforts should be encouraged across different fields, including technology-related sciences, social sciences, and humanities. Sharing knowledge and best practices can help tackle conceptual and methodological challenges specific to each discipline.
- Understanding subject-specific communication behaviour within various publication types and channels is crucial. Research on how citations (patents classic scholarly links as well as novel social media and web links) function differently in different domains can lead to improved metrics and analysis. For instance, citation indicators do also encompass patent citations. Publications may refer to patents while patents do refer to publications[19]. References in patents, both to the patent and the non-patent literature, provide a relevant diversification of citation patterns beyond the citation universe present in the publication literature.
- To address meso- and micro-level issues like co-authorship, gender, and open-access (OA) publication, data processing approaches should aim for higher granularity, providing more detailed information for accurate analysis.

17 Rafael Ball (Ed.), *Handbook Bibliometrics*. De Gruyter Saur, 2020. ISBN: 978-3-11-064227-8.
18 Wouters, Paul, Judit Bar-Ilan, Mike Thelwall, Isidro F. Aguillo, Ülle Must, Frank Havemann, Hildrun Kretschmer, et al. “Acumen Final Report.” Brussels: European Commission, 2014. http://cordis.europa.eu/result/rcn/157423_en.html
19 Ahmadpoor, Mohammad, and Benjamin F. Jones. “The Dual Frontier: Patented Inventions and Prior Scientific Advance.” *Science* 357, no. 6351 (August 11, 2017): 583–87. https://doi.org/10.1126/science.aam9527

By implementing these recommendations, the field of scientometrics can overcome challenges and continue to evolve with the changing landscape of scholarly communication.

3.2 Strengths of the original scientometric model

Despite concerns regarding the misuse of metrics (as mentioned previously and in subsequent sections), the original scientometric model had several strengths. First, it relied on unique, multidisciplinary bibliographic databases such as ISI Science Citation Index, and later, its successors – Thomson Reuters/ Clarivate's Web of Science and Elsevier's Scopus, which provide a closed but dynamic universe. While these databases are expanding, they still form a closed universe. Any links with items outside the databases, such as non-source items in references, can be filtered out and excluded from the analyses. This has allowed for standardisation and integration of quality-consistent indicators, which facilitate the comparability of scientometric results by providing a clear definition of exposure and scholarly impact. Second, since it was limited to the measurement of scholarly communication, it provided clear definitions of actors, impact, and users of information within the scholarly framework. This makes the interpretation of scientometric results more straightforward.

The availability of proprietary data products also makes it highly reproducible and documentable, as long as transparent documentation and descriptions are provided. The approach works at any level of aggregation and can be combined with peer review systems, allowing for the analysis of (inter-)national and (inter-)institutional levels, as well as individual researchers and research teams. Finally, mathematical-statistical models could be applied successfully to a variety of processes, including publication activity, citation impact, co-authorship, citation-based networks, and literature growth and evolution. While the system remains closed, it is possible to extend it to an open dynamic universe by matching databases with external sources, even if the information is incomplete.

3.3 Challenges and limitations of Scientometrics 2.0

Recently, Scientometrics 2.0 was proposed which would encompass measures of societal impact and the "broader impacts" of research, as well as "open science" - social media metrics or alternative metrics as components of a "Scientometrics 2.0" foundation[20]. To achieve this, the integration of new sources into the analyses such as bookmarking, reference managers, recommendation systems, comments on articles, news sources and policy documents, microblogging, Wikipedia, blogging, preregistrations and preprints, and social networks, as well as video and open data repositories, were recommended. This expanded scope and the use of alternative metrics could provide a more comprehensive and nuanced view of the impact of research, including the societal and economic impact beyond academia.

While the concept of Scientometrics 2.0 promises to measure the societal and broader impacts of research and covers alternative metrics, there are several challenges and limitations that need to be addressed. While new sources offer complementary possibilities, they have specific limitations when it comes to benchmarking and generalisations. Wouters and Costas (2012)[21], Sugimoto (2016)[22], and Gumpenberger et al. (2016)[23] have summarised these challenges as follows:

- Most analyses are conducted at the individual level, and the aggregation of results at higher levels is questionable, making large-scale studies difficult to conduct. This raises issues of validity, reliability, and feasibility.
- Several assumptions are not yet validated, and the rapid development of online and electronic communications poses a challenge for altmetrics to keep up with developments in scientometrics.
- More transparency and clarity in the data covered are needed. There is no clear definition of actors on both sides, so the standardisation and normalisation of measures are hardly conceivable without clarification of impact and potential biases.
- Data quality is a concern, not only for source-related reliability of input and assignments but also for the results themselves, as automated processes may produce errors and influence social media metrics.

20 Priem, Jason and Bradley M. Hemminger (2010). Scientometrics 2.0 New Metrics of Scholarly Impact on the Social Web", *First Monday* vol. 15, nr. 7, 5 July 2010. https://firstmonday.org/ojs/index.php/fm/article/download/2874/2570

21 Wouters, Paul, and Rodrigo Costas. "Users , Narcissism and Control – Tracking the Impact of Scholarly Publications in the 21 Century." Utrecht: SURF, SURFHARE, 2012. https://www.researchgate.net/publication/268271559_Users_Narcissism_and_Control-Tracking_the_Impact_of_Scholarly_Publications_in_the_21_st_Century

22 Sugimoto, Cassidy (2016). " Social Media Metrics as Indicators of Broader Impact Altmetric Indicators Do Not Measure a Broader Array of Genres than Citations. ." OECD: Organisation for Economic Co-operation and Development. https://web-archive.oecd.org/2016-10-20/418029-172%20-%20SugimotoOECDaltmetrics.pdf

23 Gumpenberger, Christian, Wolfgang Glänzel, and Juan Gorraiz. "The Ecstasy and the Agony of the Altmetric Score." Scientometrics 108, no. 2 (August 2016): 977–82. https://doi.org/10.1007/s11192-016-1991-5

- Altmetrics lacks mathematical background and proper models, which impedes the clear interpretation of indicators. Composite indicators and arbitrary construction of new 'all-in-one' indicators pose issues for interpretation and comparability, and we should avoid the temptation to create new indicators for ranking.

As far as science policy is concerned, the problem of the inappropriate use of scientometrics in science policy has been well-documented, and it ranges from uninformed use to the deliberate misuse of data. Bibliometric research is not always free from responsibility for uninformed use and misuse. Unfortunately, bibliometricians may be tempted to follow popular or populist trends to meet the expectations and demands of customers or policymakers. The advent of Scientometrics 2.0, or the next phase of bibliometric research, poses several challenges to both the discipline and its practitioners. It is essential to steer clear of the pitfalls associated with the improper use of new indicators, whose meanings and mathematical foundations are not yet fully understood or sufficiently validated[24].

24 Debackere, K., Glänzel, W., Thijs, B. (2019). Scientometrics Shaping Science Policy and vice versa, the ECOOM Case. In: Glänzel, W., Moed, H.F., Schmoch, U., Thelwall, M. (eds) Springer Handbook of Science and Technology Indicators. Springer Handbooks. Springer, Cham. https://doi.org/10.1007/978-3-030-02511-3_17

4. The limitations of peer review and score-based grant criteria

Grants for research and innovation are typically assigned based on (peer) evaluation and score-based grant criteria.[25] The practice of allocating funding based on review scores, however, has been met with considerable criticism. Peer review of grants is biased against risky projects with high potential for radical innovation and towards conservative projects with potential for short-term gains.[26] In the context of technology development, some scholars even recommend that researchers conduct "stealth innovation" projects that fly under the radar as long as possible;[27] such practices are perhaps a consequence of the apparent inability of traditional grant evaluation to value and fund research that is high-risk, high-reward. These issues become even more prevalent for projects that are close to the funding cut-off.[28] With a need for impactful, transformative innovation policy to meet societal challenges[29] and slowing progress in disruptive science and technology[30], avoiding such biases is crucial. The current system is also biased against replication research, which is often seen as less innovative.

Scholars and policymakers have started experimenting with alternate funding allocation schemes in response to these challenges. Some allow for a dimension of randomness by conducting lotteries among proposals of sufficient quality[31]. Others have devised ranking methodologies that consider uncertainty in the ranking of proposals close to the funding cut-off[32].

Several institutions, including the Volkswagen Foundation[33], the Health Research Council of New Zealand[34], the Austrian Research Fund[35], and the Swiss National Science Foundation[36], and the Research on Research Institute[37] have already set up experimental programs that include random elements for assigning transformative research funding.

The results of these experiments can be used for policy recommendations for developing next generation metrics to inform the outcomes of experimental funding programs, for designing new funding schemes aimed at transformative innovation or for redesigning existing funding lines.

4.1 Recommendations for use of metrics in funding allocation schemes to address biases and promote transformative innovation

- Develop ranking methodologies that consider uncertainty in the ranking of proposals, especially those close to the funding cut-off. By considering the uncertainty inherent in evaluations, funding agencies can avoid overlooking potentially impactful projects that may be overlooked in a traditional score-based evaluation system.

25 Liaw, Lucy, Jane E. Freedman, Lance B. Becker, Nehal N. Mehta, and Laura Liscum (2017). Peer review practices for evaluating biomedical research grants: A scientific statement from the American Heart Association. Circulation Research, 121, e9-e19. https://doi.org/10.1161/RES.0000000000000158
26 Alberts, Bruce, Marc W. Kirschner, Shirley Tilghman, and Harold Varmus (2014). Rescuing US biomedical research from its systemic flaws. Proceedings of the National Academy of Sciences of the United States of America, 111, 5773-5777. https://doi.org/10.1073/pnas.1404402111; Guthrie, Susan, Chiga Ioana, and Steven Wooding (2018). What do we know about grant peer review in the health sciences? F1000 Research, 6, 1335. https://doi.org/10.12688/f1000research.11917.2
27 Miller, Paddy, and Thomas Wedell-Wedellsborg (2013). The case for stealth innovation. Harvard Business Review, March, 91-97. https://hbr.org/2013/03/the-case-for-stealth-innovation
28 Adam, David (2019). Science funders gamble on grant lotteries, Nature, 575, 574-575. https://doi.org/10.1038/d41586-019-03572-7
29 Boon, Wouter, and Jakob Edler (2018). Demand, challenges, and innovation. Making sense of new trends in innovation policy. Science and Public Policy, 45, 4, 435-447. https://doi.org/10.1093/scipol/scy014
30 Park, Michael, Erin Leahey, and Russel J. Funk (2023). Papers and patents are becoming less disruptive over time. Nature 613, 138-144. https://doi.org/10.1038/s41586-022-05543-x
31 Fang, Ferric C., and Arturo Casadevall (2016). Research funding: the case for a modified lottery. mBIO, 7(2), e00422-16. https://doi.org/10.1128/mbio.00422-16
32 Heyard, Rachel, Manuela Ott, Georgia Salanti, and Matthias Egger (2022). Rethinking the funding line at the Swiss National Science Foundation: Bayesian ranking and lottery. Statistics and Public Policy, 9(1), 110-121. https://doi.org/10.1080/2330443X.2022.2086190
33 Volkswagen Foundation (2017). Experiment! – In search of bold research ideas. https://www.volkswagenstiftung.de/en/funding/funding-offer/experiment-search-bold-research-ideas-completed
34 Liu, Mengyao, Vernon Choy, Philip Clarke, Adrian Barnett, Tony Blakely, and Lucy Pomeroy (2020). The acceptability of using a lottery to allocate research funding: a survey of applicants. Research Integrity and Peer Review 5, 3.
35 Austrian Research Fund (2020). 1000 Ideas programme. https://www.fwf.ac.at/en/research-funding/fwf-programmes/1000-ideas-programme
36 Bieri, Marco, Katharina Roser, Rachel Heyard, and Matthias Egger (2021). Face-to-face panel meetings versus remote evaluation of fellowship applications: simulation study at the Swiss National Science Foundation. BMJ Open, 11, e047386.
37 https://researchonresearch.org/about/

- Analyse the outcomes of experimental funding programs that have already incorporated random elements in funding allocation. Institutions like the Volkswagen Foundation, Health Research Council of New Zealand, Austrian Research Fund, and Swiss National Science Foundation have set up such programs, and their results can provide valuable insights and inform the design of new funding schemes.
- Develop collaboration with funding organizations, for example via the Research on Research Institute (RORI) and by linking with the European Coalition on Reforming Research Assessment (CoARA) which aims for a more inclusive and responsible research assessment.

By implementing these recommendations and linking them to next generation metrics, funding agencies and universities can enhance the fairness, impact, and effectiveness of their allocation strategies, ultimately fostering a research ecosystem that encourages transformative innovation and addresses societal challenges more effectively.

5. Next generation metrics' complex interrelationship across multiple levels of the academic system

When developing performance indicators and research metrics, it is important to consider the complex interactions between different levels in the international academic system. Funding arrangements are organised at both the national and supra-national (European Union) level, while regular formalised assessment exercises are also organised nationally in many countries. At the same time, universities have developed strategic approaches that complement the bottom-up initiatives of academic staff with top-down branding, valorisation policies, consortium building and occasionally, even the amalgamation of universities to create larger entities. For example, the establishment of Sorbonne Université in January 2018 was a merger of Paris-Sorbonne University (Université Paris-Sorbonne) and Pierre and Marie Curie University (Université Pierre et Marie Curie). At a granular level in universities, academic groups are connected to larger international networks that are independent of national and university structures. Such networks play a critical role in shaping the future of the academic staff involved, as they govern international and national publications, conferences, and serve as pathways for future career opportunities.

5.1 Recommendations for context-aware development and implementation of NGM

The specific recommendations, based on the discussed interrelationship between the different levels of the academic system and the importance of context in next generation metrics, are as follows:

- Universities and academic institutions should adopt a holistic approach in understanding the interrelationship between different levels of the academic system. This includes recognising the circular nature of the system and the influence of international networks in research and teaching.
- Metrics should recognise and reward the contribution of teams and not just individuals.
- When developing and implementing next generation metrics, institutions must be highly attentive to the contexts in which these metrics will be extracted and applied. Metrics should be designed and adapted to suit the specific levels of the academic system, considering the diversity of interactions and complexities involved.
- Recognise the inherent complexity of interactions within the academic system. While guidelines can help, it is essential to be aware that certain dynamics may defy simple constraints. Therefore, ongoing assessment and adaptation are necessary to address evolving complexities.
- Universities must, therefore, make optimal use of the expertise of metrics specialists to consider the various dimensions of next generation metrics and how they interact with the academic system's levels. Given the limited resources of universities, pooling metrics expertise and collaboration between universities and funding agencies seems the best way forward.

By following these recommendations, universities can better navigate the interplay between different levels of the academic system and the context of next generation metrics. This will contribute to the development of a more robust and responsive framework for evaluating research and academic activities.

5.2 The push for next generation metrics

The need for next generation metrics, driven in part by the emergence of novel digital data, is primarily focused on two areas:

1. Tracking the progress of new academic policies, such as open science and reward and recognition, and assessing their levels of awareness and adoption within the academic community[38].
2. Developing new methods for evaluating academic performance, including the evolution towards team science.

At the international and national level, next generation metrics are expected to inform policy making at high levels of abstraction. They provide both quantitative and qualitative information to check compliance and detect emerging issues, such as unexpected effects of policies. The scientific performance of organisations such as the OECD, European Union, or a country can be estimated based on aggregate data and indicators.

38 Wouters, Paul, Ismael Ràfols, Alis Oancea, Shina Caroline Lynn Kamerlin, J. Britt Holbrook, and Merle Jacob. Indicator Frameworks for Fostering Open Knowledge Practices in Science and Scholarship. LU: Publications Office, 2019. https://data.europa.eu/doi/10.2777/445286

Recently, team science has been added to the academic organisational-institutional context, necessitating the development and application of metrics at yet another level of analysis.

5.3 Team science

In recent decades, science teams and larger groups have gained prominence in the scientific enterprise. They exhibit diverse membership, require deep knowledge integration, possess large sizes, face goal misalignment with other teams, have permeable boundaries, experience geographic dispersion, and exhibit high task interdependence. These features and challenges necessitate new metrics for measuring outputs, assessing contributions, and benchmarking. Expert and peer review processes further complicate matters, as goal misalignment can occur not only at the team level but also at the expert and funding levels. As a result, team science requires specific metrics, assessment methodologies, and monitoring techniques that extend beyond traditional measurements. This presents challenges related to data, indicators, and methods.

In 2015, the US National Academy of Sciences (NAS) provided definitions for team science, science teams, and larger groups. Team science refers to scientific collaboration conducted by multiple individuals in an interdependent manner, including both small teams and larger groups. Science teams typically consist of two to ten individuals, while larger groups involve more than ten individuals and often include multiple smaller science teams. These larger groups have a differentiated division of labour and an integrated structure to coordinate the smaller teams, and they are referred to as organisations in the social sciences. The effectiveness of a team, also known as team performance, is measured by its capacity to achieve goals and objectives. This leads to improved outcomes for team members and outcomes influenced by the team, such as new research findings or translational applications.[39]

With the growth of specialised knowledge in various fields, individual researchers find it more challenging to integrate knowledge from diverse backgrounds. Therefore, collaborating with colleagues who possess complementary expertise has become a preferred approach to investigate questions beyond their narrow specialisation. Science teams and larger groups vary in their degree of integrating knowledge from different disciplines or professions to achieve scientific and translational goals. Unidisciplinary research relies on a single discipline, while multidisciplinary research involves separate contributions from each discipline. Interdisciplinary research integrates information, data, techniques, and perspectives from multiple disciplines, addressing complex scientific and societal problems. Transdisciplinary research not only integrates but also transcends disciplinary approaches.[40]

39 National Academy of Sciences, Engineering and Medicine, 2015, Enhancing the effectiveness of team science.
40 Hall, K.L., Vogel, A.L., Croyle, R.T. (eds.), 2020, Strategies for team science success, Springer.

6. The metric 'trickle-down' challenge

One of the main issues in the discussion about the misuse of performance metrics such as the Journal Impact Factor and the H-index is their use at inappropriate levels of analysis and implementation. This problem is not solely due to a misunderstanding but has its origins in specific organisational dynamics, both of which reflect broader issues in research culture. As mentioned previously in section 5, understanding the interplay between these different units and levels is crucial.

6.1 Factors contributing to the trickling down of indicators

Several partially independent factors have contributed to the "trickling down effect" of indicators. The current lack of easily available context sensitive metrics leads to the use of those indicators that are available, even if they are inappropriate. This is exacerbated by the dynamics of competition for reputation among researchers, which can prioritise indicators that have become symbols of reputation. This is enhanced by the influence of global university rankings and their reliance on a specific set of indicators and metrics. Moreover, most universities do not have the capacity to accurately monitor their policies and performance with state-of-the-art data and measurement technologies. This means that they cannot extract meaningful and context sensitive indicators from their own data sets. The historical development of science policy (see earlier in this report) has also contributed by creating strong links between a narrow set of publication, citation, and funding indicators to performance measurement and evaluation, both in universities and by funding agencies. It has for example resulted in the standardization of Curricula Vitae and personal web pages of researchers that are mainly a list of publications, grants and awards[41]. In addition, annual performance interviews are still lacking standards in most countries. The challenge of translating qualitative judgements and discussions about individual results into summary results at higher levels of aggregation has stimulated the trickling down effect. This is generally done only once every few years in national cycles of formal research evaluations, such as in the UK and the Netherlands. The perception that assessment based on qualitative indicators (e.g., narrative CVs) is less objective and is more time consuming is a barrier to change of the evaluation system. This is especially problematic given the increased time pressures and workloads that most academics face.

6.2 Recommendations to mitigate misuse of performance metrics in academic evaluation

All of the above factors can contribute to the misuse or overuse of performance metrics and indicators at the wrong level of analysis or implementation. It is important for universities to be aware of these potential pitfalls and to take steps to mitigate them, such as providing more nuanced metrics at lower levels of aggregation and avoiding funding models that rely too heavily on specific indicators.

It should also be noted that such factors could signpost possible solutions for next generation metrics, for example:

- Improving and standardising parts of the university registration systems can ensure that relevant data is available for the generation of metrics, and that the metrics generated are reliable and consistent.
- Not linking funding and allocation to performance metrics at higher levels of aggregation, but only to the mission and goals at the same level can help prevent the trickling down of inappropriate metrics and avoid negative consequences.
- Standardising performance interviews can ensure that qualitative data is consistently collected and can be used to supplement quantitative metrics.
- Using computational linguistic tools to extract relevant metrics from annual performance interviews (or not!).
- Adapting the use of next-generation policy and performance metrics to align with the relevant evaluation cycle and restricting their routine application can help ensure their appropriate utilisation and prevent misuse.
- Diversifying and developing new standards for academic CVs can help promote a broader range of metrics that better reflect the diversity of academic achievements and contributions.[42]
- Changing the use of performance metrics by funding agencies can help avoid negative consequences and ensure that metrics are used appropriately.[43]
- Developing more responsible university policies towards global university rankings can help avoid the undue influence of rankings on university policies and practices.

41 Woolston, Chris. "Time to Rethink the Scientific CV." *Nature* 604, no. 7904 (April 5, 2022): 203–5. https://doi.org/10.1038/d41586-022-00928-4
42 https://www.leidenmadtrics.nl/articles/narrative-cvs-a-new-challenge-and-research-agenda
43 https://researchonresearch.org/

- Promoting the use of a diverse portfolio of metrics as new symbols of reputation can help avoid the negative consequences of over-reliance on a limited set of metrics.

The Coalition for Advancing Research Assessment[44] is an example of an initiative addressing the above recommendations at an international (European) level.

44 CoARA - Coalition for Advancing Research Assessment

7. Next generation metrics and university rankings

University rankings compare and assess higher education institutions based on various criteria to provide insights into their quality and reputation. Ranking systems such as the Times Higher Education[45], QS World University Rankings[46], ARWU[47] and CWTS Leiden Ranking[48], use a range of different metrics from academic reputation, research output and international collaboration, to student-staff ratios, research income and employment outcomes to generate their rankings tables. These rankings agencies generally have a commercial focus that aims to generate revenue through various means. This includes gathering data from universities and other sources to assess performance indicators, developing methodologies for evaluation, analysing the collected data, and publishing rankings through websites and reports. These entities also generate revenue through licensing, advertising, sponsorships, and organising conferences and events.[49]

7.1 The importance and influence of rankings

In an ideal world, university rankings based on one-size fits all indicators should not matter. The nature of the global higher education landscape at present, however, means that rankings matter much more than we would like. As a result, universities feel obliged to engage with them to protect the mission of the university. While rankings are a controversial metric, they can offer a snapshot of a university's international reputation and profile. They can also be valuable for prospective overseas students and early career researchers when deciding which universities to apply to. Rankings can also be used by universities to help build strategic collaborations and inform analyses of where they are as a university, and where they want to go. Used strategically, rankings can also be a powerful lobbying tool when making the case to a national government for increased investment in the third-level sector. Policymakers also often refer to these rankings to assess the performance and competitiveness of their country's higher education institutions on a global scale, to develop internationalisation strategies, educational and funding policies and to benchmark and incentivise their national higher education institutions. In India, for example, the University Grants Commission (UGC) has initiated measures to internationalise the higher education system. As part of these schemes, the UGC requires foreign partners of Indian universities to be ranked in the top 500 of THE or ARWU rankings.[50]

In the last two decades, university rankings have also become part of the culture of research evaluation in many countries including the UK, Australia, China, Germany, and the Netherlands. While the results of research assessment are rarely ordered in a hierarchical manner, publication of their results by the media or other organisations has often led to the production of a 'league table' of HEIs. This practice has facilitated the restructuring of the higher education system and has led to a growing convergence between assessment and rankings.[51] University rankings have also been used in unexpected and anomalous contexts, such as being incorporated into immigration schemes or being considered as trusted metrics that influence national or global reports, with various countries implementing measures that incorporate university rankings into their policies and programs. The UK, for example, recently launched a visa scheme specifically for graduates from the world's top 50 non-UK universities, with eligibility criteria requiring universities to appear in the top 50 rankings of at least two major ranking systems.[52] Similarly, the Netherlands offers the 'Dutch Highly Educated Migrant Visa', allowing recent graduates from internationally recognised universities to live in the country for one year to find employment as Highly Skilled Migrants. To qualify, the university must appear in the top 200 of general ranking lists or ranking lists by faculty or subject field in the Times Higher Education (THE), Quacquarelli Symonds (QS), or Academic Ranking of World Universities (ARWU).[53]

University rankings are not only influential in immigration and educational policies but also impact other areas. The International Institute for Management Development (IMD) produces an annual World Competitiveness Ranking, which serves as a reference point for a country's competitiveness. This ranking is utilised by

45 https://www.timeshighereducation.com/world-university-rankings/2023/world-ranking
46 https://www.topuniversities.com/university-rankings/world-university-rankings/2024
47 https://www.shanghairanking.com/rankings/arwu/2022
48 https://www.leidenranking.com/
49 https://www.timeshighereducation.com/events
50 www.ugc.gov.in/pdfnews/9214094_Draft-Setting-up-and-Operation-of-Campuses-of-Foreign-Higher-Educational-Institutions-in-India-Regulations-2023.pdf
51 Hazelkorn, E. (2009). Pros and Cons of Research Assessment: Lessons From Rankings. World Science, 2009. https://arrow.tudublin.ie/cserrep/24/
52 https://www.gov.uk/high-potential-individual-visa
53 https://immigration-portal.ec.europa.eu/netherlands-highly-qualified-worker_en

corporations as part of the foreign direct investment process, with the Times Higher Education university ranking being one of the metrics used in the calculations for this report.[54] In addition, various scholarship programs also consider university rankings. The Korea Foundation for Advanced Studies (KFAS) provides scholarships to Korean students admitted to prestigious universities abroad for graduate studies. To be eligible, students must be enrolled at a top 100 university.[55] Similarly, the Albanian Government Excellence Fund supports Albanian students admitted to the top 15 universities ranked by the Times Higher Education World University Ranking, both overall and/or in their chosen subject of study.[56]

7.2 Criticisms and limitations of rankings

To be truly beneficial, therefore, there is a need to understand the structure, measurements, and limitations of various rankings. Criticisms have arisen due to the necessity of comprehensively understanding the nuances of rankings and their impact on decision-making processes. This criticism encompasses a wide range of issues including conceptual concerns, such as the emphasis on competition over collaboration, the self-perpetuating nature of ranking positions leading to a Matthew effect, and the focus on output quantity rather than value-added contributions.[57] Additionally, the choices of data and variables used in rankings are criticised for providing a limited view of universities' activities and values. In the Netherlands, an expert group advising the Board of Universities (UNL) has reached a similar conclusion regarding global university rankings. Dutch universities will now adopt a more critical approach towards these rankings, as they believe that the league tables suffer from methodological limitations. They argue that the current rankings focus too much on one aspect of research achievements, primarily relying on the number of publications and citations and feel that this narrow perspective clashes with the Recognition & Rewards program, which prioritises quality over quantity in evaluating the academic performance of Dutch universities.[58] Rankings are also susceptible to the streetlight effect by relying on easily available data and favouring data sources that disadvantage non-English speaking and Global South nations.[59]

7.3 Reforming university rankings

Methodologies employed in rankings have also been criticised by a growing number of researchers for being methodologically flawed, lacking transparency, and oversimplifying the complexity of universities into composite numbers.[13, 60] Concerns have also been raised regarding the lack of open data, the commercial nature of most rankings without independent oversight, and the increasing reliance on the rankings industry. In response to these growing concerns, the INORMS project[61], aimed at rethinking university rankings, developed a methodology to compare ranking systems[62] based on rigour, governance, measurement of relevant criteria, and transparency. The project concluded that all the rankings included in the comparison fell short on these criteria, particularly in terms of "measuring what matters to the communities they were ranking". Such criticisms highlight the limitations and shortcomings of university rankings, emphasising the need for comprehensive reform to address issues related to methodology, data transparency, community relevance, and oversight. University rankings are also addressed in various ways in a range of LERU reports, including the LERU OS roadmap[63] and LERU open science note[64], as well as in other previous recommendations about the responsible use of metrics (see **Appendix II**)

There are also a range of statements and declarations on research assessment and research metrics that lack coherence in their approach to university rankings. DORA (the San Francisco Declaration on Research Assessment)[65] focuses on the assessment of individual researchers, and therefore does not address university rankings directly. The Metric Tide report[66] recommends that institutions at least explain why they use rankings and to what ends, and it calls for ranking suppliers to provide transparency on their data and methods.

54 https://www.imd.org/centers/wcc/world-competitiveness-center/rankings/world-competitiveness-ranking/
55 https://eng.kfas.or.kr/theme/kfaschanel/intl_scholarship_5.php
56 https://www.privacyshield.gov/article?id=Albania-Education-and-Training
57 Kaidesoja, T. (2022). A theoretical framework for explaining the paradox of university rankings. Social Science Information, 61(1), 128–153. https://doi.org/10.1177/05390184221079470
58 Universiteiten van Nederland (UNL), DANS Data Station Social Sciences and Humanities, V1, 2023, "Ranking the university", https://doi.org/10.17026/SS/XGDX0H
59 Nassiri-Ansari, T., & McCoy, D. (2023). World-class universities? interrogating the biases and coloniality of global university rankings. https://collections.unu.edu/eserv/UNU:9082/UNI-Ranking-paper_February-2023.pdf
60 Soh, K. (2017). The seven deadly sins of world university ranking: A summary from several papers. Journal of Higher Education Policy and Management, 39(1), 104-115. https://doi.org/10.1080/1360080X.2016.1254431
61 INORMS. Rethinking Global University Rankings. https://inorms.net/rethinking-global-university-rankings-3/
62 Gadd, E., Holmes, R., & Shearer, J. (2021). Developing a method for evaluating global university rankings. Scholarly Assessment Reports, 3(1). https://doi.org/10.29024/sar.31
63 Open Science and its role in universities: a roadmap for cultural change. Paul Ayris, Alea López de San Romàn, Katrien Maes, Ignasi Labastida. LERU Advice Paper no.24, May 2018. https://www.leru.org/publications/open-science-and-its-role-in-universities-a-roadmap-for-cultural-change
64 Implementing Open Science, Paul Ayris. LERU Note, January 2021. https://www.leru.org/publications/implementing-open-science
65 https://sfdora.org/read/
66 Wilsdon, J., et al. (2015). The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management. https://doi.org/10.13140/RG.2.1.4929.1363

Many of the issues raised by the Metric Tide report regarding data and methodology in measuring research output also apply to university rankings. Similarly, the Leiden Manifesto[67] highlights problems with inaccurate data, arbitrary criteria, lack of transparency, and other methodological issues related to research assessment, including university rankings. Although the EU NextGen Metrics Report[68] does not explicitly address university rankings, the alternative metrics discussed in the report could potentially be incorporated into university ranking systems.

The Hong Kong Principles[69] do not specifically address university rankings, but their recommendations on using a range and diversity of assessment criteria can be applied to all types of assessment, including university rankings. Similarly, INORMS/SCOPE is a framework for evaluation that can be applied to ranking systems. The Coalition for Advancing Research Assessment (CoARA)[70], signed by many LERU member institutions, explicitly addresses rankings, and advises against ranking research organisations in research assessments to prevent the influence of rankings on the assessment of individual researchers and research groups. It also emphasises the importance of maintaining control over ranking methodologies and data. In addition, More Than Our Rank[71] is a statement dedicated solely to university rankings, but its focus is openly rallying support rather than providing guidance on how to proceed. This approach can inspire collaborative efforts to develop new approaches to university rankings by universities themselves (e.g. in the Netherlands[72]). More details on some of the reports, frameworks and guidelines mentioned above can be found in **Appendix II**.

7.4 Addressing alternative metrics and adjustments

Some rankings, however, are striving to incorporate more balanced metrics in their evaluation processes, or they adjust their metrics based on the changing priorities of the institutions they evaluate. Many have for instance introduced discipline-specific ranking. Some have introduced new variables such as the contribution to sustainable development goals (SDGs), open access or gender diversity. An example is the CWTS Leiden Ranking, which utilises indicators like scientific impact, gender diversity, open access, and collaboration as alternative metrics.[73] These metrics provide a broader and more comprehensive perspective on research performance, acknowledging factors beyond traditional measures and offer a more nuanced and inclusive assessment of universities and their contributions. This shift has also been seen in the recent QS rankings, where there was a de-emphasis on survey-based metrics and an inclusion of additional metrics based on international collaborations, graduate outcomes, and sustainability. The QS rankings methodology, however, still places a considerable amount of emphasis on traditional scholarly metrics and the more fundamental issues are not being addressed.[74]

While the current university rankings' limitations are reason to avoid promoting or using them in research assessment and assessment of university policies, this does not exclude the use of quantitative data for research activities once the data provides a balanced and valid view on trends and patterns. Universities could reconsider how they use the data, rather than the ranking lists. For instance, data could be used in a more nuanced and targeted way, looking at the separate variables instead of composite metrics and at the ratio rather than the ordinal measurement level. When comparing it may be interesting to compare not so much with supposed competitors but compare one's own institution, team or programme over time or compare with the one's own goals. However, generating and using data is something that can also, or preferably, be done outside the current commercial ranking provider services.

7.5 Recommendations for university rankings and their interpretation

- While alternative approaches, like the CWTS Leiden ranking, could provide more comprehensive assessments and serve as potential alternatives to traditional ranking systems, it is still unlikely that traditional ranking systems will cease to exist. Instead, there is a need to develop guidelines for interpreting rankings, including approaches that address emerging criteria like sustainability and open science.
- Following the Metric Tide report's advice, institutions should be transparent about their use of rankings, explaining the purpose and intentions behind their application. Ranking suppliers must also disclose their data and methodologies to ensure accountability and fairness.
- Although the Hong Kong Principles do not directly address university rankings, their principles of diverse assessment criteria can be applied to improve the fairness and inclusivity of ranking systems.

67 Leiden Manifesto for Research Metrics. 2015. http://www.leidenmanifesto.org/
68 European Commission, Directorate-General for Research and Innovation, Peters, I., Frodeman, R., Wilsdon, J., et al., Next-generation metrics : responsible metrics and evaluation for open science, Publications Office, 2017, https://data.europa.eu/doi/10.2777/337729
69 Moher D, Bouter L, Kleinert S, Glasziou P, Sham MH, Barbour V, et al. (2020) The Hong Kong Principles for assessing researchers: Fostering research integrity. PLoS Biol 18(7): e3000737. https://doi.org/10.1371/journal.pbio.3000737
70 https://coara.eu/agreement/the-agreement-full-text/
71 https://inorms.net/more-than-our-rank/
72 https://www.universiteitenvannederland.nl/en/ranking-the-university
73 https://www.leidenranking.com/information/indicators
74 https://support.qs.com/hc/en-gb/articles/4405955370898-QS-World-University-Rankings

- Policymakers and institutions should utilise the INORMS/SCOPE framework for evaluation to assess ranking systems and ensure they align with their goals and priorities. The Coalition for Advancing Research Assessment (CoARA) advises against ranking research organisations in research assessments to avoid undue influence on individual researchers and research groups, underlining the importance of maintaining control over ranking methodologies and data.
- Ranking systems should proactively address fundamental issues in their methodologies, considering evolving priorities such as discipline-specific ranking and factors like sustainable development goals, open access, and gender diversity.
- While recognising the limitations of current university rankings, institutions can still use quantitative data for research activities if balanced and valid. They should explore nuanced approaches like analysing separate variables rather than composite metrics and focusing on ratios instead of ordinal measurement levels. Comparisons should extend beyond competing institutions to evaluate one's own progress and alignment with internal goals.
- To mitigate the negative impact of competitiveness in rankings, data should be presented through grouped metrics rather than relying solely on composite ranking positions. Moreover, institutions must responsibly inform prospective students and researchers about the limitations of rankings. In addition, they should educate journalists, governments, and industry stakeholders about the potential risks of using rankings as the sole basis for funding or policy decisions, and (mis)communicating university ‘prestige’ to the public.
- To enhance the inclusivity and credibility of rankings, it is essential to include a range of 'new' open scholarship indicators and promote greater transparency in the data and methodologies used.

8. New metrics and the emergence of related ethical and technical challenges

The academic community has used a wide range of traditional metrics (citation metrics, H-index, publication in high-impact-factor journals) to measure various aspects of research performance and impact. Such metrics have also become interwoven into university processes around career development, hiring, retention and promotion. They have also driven decision-making around competitive grant funding. While traditional measures of scientific relevance still have great importance, the last decade has seen a move towards more DORA-compliant metrics.[75] This is a worldwide initiative to improve the way in which the output of scientific research is evaluated by funding agencies, academic institutions, and other parties.[76] This shift to next-generation metrics, however, potentially creates ethical and technical challenges for the citing, referencing and recognition of scholarly outputs.

Open access metrics refer to the methods used to evaluate the impact and reach of scholarly articles that are freely available online to the public, without any cost or access barriers. Output-level metrics may include the number of times an article (or dataset, code, method, or other output) has been downloaded from a repository, preprint server, or publisher's website; the number of views therein; and the number of citations attracted. Higher-level metrics provide information about outputs on an institutional or higher level. These include the percentage of an institution's publications that are open access; the percentage of publications that are linked to open datasets, methods, and/or or code; and the percentage of clinical trials or other studies that have been preregistered. Examples of such metrics are those reported by the Charité Berlin dashboard[77]. These metrics are designed to measure the visibility, usage, and impact of open access outputs and can help researchers, publishers, and institutions to assess the effectiveness of their open access policies and strategies. The advantages of open access outputs (publications, FAIR data sets, code, and methods) for authors are that they provide more exposure for a research work, make it easier for practitioners to apply research findings, yield higher citation rates, create visibility for policymakers so that a research work can more effectively influence policy, provide the public with access to research findings, meet compliance regulations for many grant funders, and, critically, can improve research reproducibility and integrity by increasing transparency. Metrics about such outputs can therefore provide valuable insights into their impact and reach and can help researchers and institutions make informed decisions about publishing and dissemination strategies.

From an ethical standpoint, however, the use of new units of analysis may raise questions around fairness and inclusivity. For example, if a new metric heavily relies on social media data, it may disadvantage individuals or groups who do not have a strong online presence. Metrics often also rely on personal data, such as author names and publication histories, which can raise privacy concerns. Ensuring that metrics are calculated in a way that protects the privacy of individuals while still providing useful information can be a challenge. Next-generation metrics may also be subject to equality, diversity and inclusivity biases. For instance, established researchers may have developed their career and academic standing based on traditional metrics. The move to new format metrics, therefore, may create a bias towards this cohort of academics. More broadly, utilising new metrics to evaluate academic performance creates the risk of bias based on gender, race, or other characteristics. It should also be stated that this is also the case for traditional metrics. For example, the h-index does not include corrections for career length, which means that this indicator disfavours younger researchers.[78] Impact factors are also biased towards a small subset of journals and disciplines and can perpetuate existing inequalities and fail to reflect the diversity of research outputs and contributors. In the move towards next generation metrics, however, cognisance can be given to these ethical issues. This will ensure that, from their inception, these new metrics are not only responsible[79], but are also fair and unbiased and that, in as much as possible, all researchers are being accurately represented.

From a technical standpoint, the use of new units of analysis may also raise questions around how to properly cite and reference the data. Altmetrics is a term to describe web-based metrics for the impact of publications and other scholarly material by using data from social media platforms (e.g. Twitter or Mendeley).[80]

75 San Francisco Declaration on Research Assessment (2013)
76 B. Pulverer. Impact fact-or fiction? EMBO J, 32 (2013), pp. 1651-1652
77 https://quest-dashboard.charite.de/#tabStart
78 Ingrid Bauer, David Bohmert, Alexandra Czernecka, et al. (2020). Next Generation Metrics. Zenodo. https://doi.org/10.5281/zenodo.3874801
79 Wilsdon, J., et al. (2015). The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management. https://doi.org/10.13140/RG.2.1.4929.1363
80 Bornmann, L. (2014). Do altmetrics point to the broader impact of research? An overview of benefits and disadvantages of altmetrics, Journal of Informetrics, 8 (4), 895-903, https://doi.org/10.1016/j.joi.2014.09.005

Such metrics provide new ways of approaching, measuring and providing evidence for impact.[81] These metrics are challenging because they require new data sources and methodologies, and their meaning and interpretation are still evolving.

Such metrics may also rely on machine learning algorithms or other proprietary technologies, which may be difficult for researchers to access the underlying data or reproduce the results. This can create challenges for peer review and for building on the findings of previous research. In addition, new metrics such as usage metrics or altmetrics can be easy to manipulate or inflate and are not straightforward to collect or reproduce. Recent controversies in the scientific community further indicate the need for replication of research designs and their empirical outcomes. Lack of replicability may lead to paper retractions and damages the reputation of science and its practitioners. Therefore, it is advised to develop, design, and deploy (novel) replication metrics as a relevant subset of the emerging portfolio of next generation metrics[82, 83].

In collaborative research, issues of attribution may arise, making it challenging to accurately credit all contributors, especially in large research consortia where quantifying individual contributions can be difficult. The advent of AI tools, driven by the increasing availability of vast datasets and enhanced computing power, is poised to significantly impact learning, research, and scholarly communication processes. This development presents both challenges and opportunities that universities must gradually and continuously understand and master.

AI will enable new inventive methods, leading to the generation of researchable hypotheses and even the involvement of robot scientists alongside human counterparts. This evolution holds the potential to enhance both research productivity and replicability. However, it also introduces challenges related to biases, equity, and potential discriminatory effects. In the era of disruptive technologies and methodologies, such as AI, the ethical dimension deserves unwavering attention from academic and policy communities[84].

8.1 Recommendations for responsible use of next generation metrics in research evaluation

In order to address ethical and technical challenges related to new (generation) metrics, the following is recommended:

- Universities and academic institutions should use clear and comprehensive standardised guidelines and standards for citing and referencing new units of analysis. This includes creating new citation formats and implementing registries to track and share data sources effectively.
- When developing new metrics, it is essential to prioritise ethical considerations. Metrics should be designed and used in ways that respect data privacy, intellectual property, and research integrity. Institutions must ensure that researchers are aware of the ethical implications and are trained to use these metrics responsibly and transparently.
- Provide training and education to researchers on how to utilise next generation metrics responsibly. Researchers should be familiar with the appropriate context and limitations of these metrics to make informed decisions and interpretations.
- While next generation metrics can be valuable tools in evaluating research performance and impact, they should not be the sole basis for evaluation. Universities should adopt a holistic approach that considers other factors, such as the quality and originality of research, real-world impact, and an academic's contributions to their field.
- The recent emergence of artificial intelligence tools presents a new realm of ethical and technical challenges and dilemmas. Databases employing AI for sentiment analysis of citations, such as Semantic Scholar and Scite, have been operational for several years, and there's an expectation of a significant increase in AI-based services and resources in the future. Tools like ChatGPT have swiftly become standard tools, even among researchers. While this development is too recent to be reliably documented, anecdotal evidence suggests that authors are already using these tools for creating their narrative CVs, research proposals, and literature reviews. However, incorporating this into review processes remains unclear. The reverse situation is equally uncharted territory: how to harness the analytical potential of AI-based textual and numerical tools. A proactive exploration of this new landscape is advisable, as it remains uncharted territory, and understanding it as soon as possible is prudent[85].

81 Adie, E. (2014). Alternative mainstream adoption. Information Professional , 23 (4), 349–351 https://doi.org/10.3145/epi.2014.jul.01
82 Youyou, Wu, Yang Yang, and Brian Uzzi. “A Discipline-Wide Investigation of the Replicability of Psychology Papers over the Past Two Decades.” *Proceedings of the National Academy of Sciences* 120, no. 6 (February 7, 2023): e2208863120. https://doi.org/10.1073/pnas.2208863120
83 Muradchanian, Jasmine, Rink Hoekstra, Henk Kiers, and Don van Ravenzwaaij. “How Best to Quantify Replication Success? A Simulation Study on the Comparison of Replication Success Metrics.” *Royal Society Open Science* 8, no. 5 (May 19, 2021): 201697. https://doi.org/10.1098/rsos.201697
84 Ise, Holly. “Abstracts Written by ChatGPT Fool Scientists.” *Nature (London)* 613, no. 7944 (2023): 423–423. https://doi.org/10.1038/d41586-023-00056-7
85 Kendall, Graham, and Jaime A. Teixeira da Silva. “Risks of Abuse of Large Language Models, like ChatGPT, in Scientific Publishing: Authorship, Predatory Publishing, and Paper Mills.” *Learned Publishing* n/a, no. n/a. Accessed November 9, 2023. https://doi.org/10.1002/leap.1578

9. Current practices and policies on Open Science at LERU universities

9.1 Rationale for surveying Open Science policies and monitoring among LERU members

The preceding sections highlight the multifaceted nature of Open Science, along with the policies and metrics that can be employed to promote it. Due to this complexity, LERU members conducted a survey in early 2023 to evaluate the policies and monitoring mechanisms employed in support of Open Science at each LERU institution. The survey responses were used to identify policies and monitoring at both institutional and national levels, and to help members reflect on their own progress on the eight pillars of open science[86]. The survey asked the following five questions:

1. What is the actual state of funding and resources available for Open Science at your institution? To what extent do you expect development in the short term? (*Also think of relevant resources coming from national and European organizations, such as the European Open Science Cloud*).
2. Are there any existing national-level policies on Open Science? Please give your own assessment and provide URLs to relevant documentation if possible.
3. Is there any existing national-level monitoring of policies on Open Science?
4. Is there any existing national-level assessment regarding Open Science (e.g. REF in the UK, OS monitoring at national level in the Netherlands and Finland)?
5. Are there dedicated staff working at implementing the OS policies in your institution? Do institutions in your country share their practices to learn from each other? Please elaborate.[87]

For questions 2-4, respondents were requested to provide their own assessments and, if feasible, URLs to the relevant documentation. In cases where certain LERU institutions did not initially respond to the survey, the necessary information was gathered through online searches and/or follow-up emails to LERU representatives from those institutions. Finally, we consulted the summary of national open science policies provided by The Council for National Open Science Coordination (CoNOSC)[88] to validate and supplement our findings. Additional survey particulars, encompassing the responses from each institution, are available on the Open Science Framework[89]. These findings were subsequently employed to outline the present status of Open Science policies and monitoring across LERU member institutions and countries, as presented in **Table 1** opposite.

9.2 Summary table for the current state of Open Science in LERU universities and countries

In **Table 1**, countries are organised alphabetically. Color-coding is utilised to indicate the status of policies and monitoring:

- Established (Green cells)
- Not yet established but with plans in place (Orange cells)
- Not established with no agreed plans (Red cells)

For further comparative purposes, additional details are presented for Open Science policies and monitoring at a European level and in the USA, with information collected from online searches.

9.3 Notable examples of Open Science policies and monitoring among LERU members (and beyond)

The results of the survey show a wide range of activities and support related to open science among LERU universities and the countries in which they are based. It is evident that various universities and countries are at varying stages of development, with some having more mature systems in place to support and monitor open science. This variability may stem from disparities in the specific needs that each university or country aims to address, and their perception of how open science can fulfil these needs. Additionally, it could be influenced by unequal access to the necessary resources for supporting and monitoring open science.

86 https://www.leru.org/publications/open-science-and-its-role-in-universities-a-roadmap-for-cultural-change
87 Information on Open Science roadmaps from LERU members is currently being collected based on the 8 pillars (see third column of Table 1).
88 https://conosc.org/os-policies/
89 Raw survey data are available on OSF at https://doi.org/10.17605/OSF.IO/NYE3P; this includes a description of how the survey raw data were used to generate the data in Table 1.

Table 1. Summary of the Current State of Open Science in LERU Universities and Member Countries

Country	National policy	Monitoring of national policy	Monitoring of OS	LERU Institutions
Belgium	International Co-operation Commission (ICC); Federal Co-operation Commission (FCC)[90]; Flemish Open Science Board (FOSB)[91]; Flemish Research Information Space (FRIS)[92]; Open Access in Belgium[93]; there are Open Access policies from the Fund for Scientific Research (F.R.S.-FNRS)[94] and the Flemish Research Foundation (FWO)[95]; the Belgian Science Policy Office (BELSPO) has mandates for Open Access to Publications and Research Data[96]	Annual reporting of (FOSB) Open Science policy at institutional level to Flemish government	Via the Flemish Open Science Board (FOSB)[97]: monitoring (with custom-build KPI's) evolution in the OS research behavior about 4 Open Science KPI's (ORCID, DMP, Open Access, open data)	KU Leuven, Policies: Open Access policy[98], Research data management policy[99], Responsible Use of Metrics[100], Rewards & incentives[101]
Denmark	National policies on OA[102] and FAIR RDM[103], and a code of conduct for Research Integrity[104].	No monitoring of policies	Danish Open Access Indicator[105] (which monitors degree only of OA, not of FAIR data)	University of Copenhagen
Finland	National Declaration for Open Science and Research (2020-2025)[106]	Monitoring of OS policies linked to bi-annual national OS monitoring	National OS monitoring[107]	University of Helsinki OS program[108]

90 https://www.belspo.be/belspo/coordination/addgrp.asp?l=fr&group=CFS-CIS%20Open%20Science
91 https://www.frdn.be/
92 https://researchportal.be/en
93 https://openaccess.be/open-access-in-belgium/
94 https://www.frs-fnrs.be/docs/Reglement_OPEN_ACCESS_EN.pdf
95 https://www.fwo.be/en/the-fwo/research-policy/open-access/
96 https://www.belspo.be/belspo/openscience/index_en.stm
97 https://www.frdn.be/about-us/policy-in-flanders-and-belgium/#the-flemish-open-science-board-fosb
98 https://www.kuleuven.be/open-science/what-is-open-science/scholarly-publishing-and-open-access/open-access-kuleuven/open-access-kuleuven#Deposit-obligation
99 https://www.kuleuven.be/rdm/en/policy
100 https://research.kuleuven.be/en/policy-figures/responsible-use-of-metrics
101 https://eur03.safelinks.protection.outlook.com/?url=https%3A%2F%2Fwww.kuleuven.be%2Fopen-science%2Fwhat-is-open-science%2Frewards-and-incentives%2Frewards-and-incentives&data=05%7C02%7Cp.f.wouters%40cwts.leidenuniv.nl%7Cd459da1d31284c88e76408dc4ca47a1e%7Cca2a7f76dbd74ec091086b3d524fb7c8%7C0%7C0%7C638469516514304777%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C0%7C%7C%7C&sdata=69aRYluNhyJdgAUSDRWNlGWQvAD%2FMPg8qXS%2F2K%2BoVRo%3D&reserved=0
102 https://ufm.dk/en/research-and-innovation/cooperation-between-research-and-innovation/open-access/Publications/?set_language=en
103 https://www.deic.dk/en/data-management/national-cooperation/National-strategy-for-FAIR-research-data-management
104 https://ufm.dk/en/publications/2014/the-danish-code-of-conduct-for-research-integrity
105 https://oaindikator.dk/en/
106 https://avointiede.fi/sites/default/files/2020-02/declaration2020_0.pdf
107 https://avointiede.fi/en/policies-materials/monitoring
108 https://www.helsinki.fi/en/research/research-integrity/open-science

Country	National policy	Monitoring of national policy	Monitoring of OS	LERU Institutions
France	- Second National plan for Open Science (2021-2024)[109], organized in thematic colleges, including a National Fund for Open Science[110] - National platform for open publications and open data (Recherche Data Gouv[111]). - Committee for Open Science[112]	- Open Science Barometer[113] - Sharing at national level (Recherche Data Gouv[114]); shared vision of strategies in HEIs upcoming.	- French Open Science Monitor[115] provides a monitor for OA publications at each French institution. - Some monitoring of open data through Recherche Data Gouv[116]. - OS is not yet at the heart of research evaluation.	**Sorbonne University**, Policies: Various OS policies available[117]. **Université Paris-Saclay**, Policy: OS policy[118] **University of Strasbourg**, Policy: OS policy[119]
Germany	Open Science policies mostly exist at the research-organisation level[120], not at the national level. The DFG (Germany's major research funding institution) has a position on Open Science[121]. The Federal Ministry of Education and Research has strategies for Open Access[122] and Open Educational Resources[123] (both in German only). The National Research Data Infrastructure Germany (NFDI) has been focussing on open/FAIR data[124].	No monitoring of policies	Charité Berlin has a dashboard for OS monitoring[125], and a specific dashboard for open access monitoring[126].	**University of Freiburg**, Policy: OS policy[127]. **Universität Heidelberg**, Policy: OA policy[128]. **Ludwig-Maximilians-Universität München**, OS resources[129] Toolbox[130]

109 https://www.ouvrirlascience.fr/second-national-plan-for-open-science-2021-2024/
110 https://www.ouvrirlascience.fr/national-fund-for-open-science/
111 https://recherche.data.gouv.fr/en
112 https://www.ouvrirlascience.fr/comite-en/
113 https://barometredelascienceouverte.esr.gouv.fr/
114 https://recherche.data.gouv.fr/en
115 https://frenchopensciencemonitor.esr.gouv.fr/
116 https://recherche.data.gouv.fr/en
117 https://www.sorbonne-universite.fr/en/commitments-open-science
118 https://www.universite-paris-saclay.fr/sites/default/files/2022-10/EXE_SCIENCE_OUVERTE_WEB_UK.pdf
119 https://scienceouverte.unistra.fr/en/strategy/open-science-policy-of-the-university-of-strasbourg
120 https://conosc.org/os-policies/#page-content
121 https://doi.org/10.5281/zenodo.7194537
122 https://www.bmbf.de/SharedDocs/Publikationen/de/bmbf/1/24102_Open_Access_in_Deutschland.pdf;jsessionid=6C9A23CEF2B18CE438B4C2B10F5D325C.live092?__blob=publicationFile&v=5
123 https://www.bmbf.de/SharedDocs/Publikationen/de/bmbf/3/691288_OER-Strategie.pdf?__blob=publicationFile&v=4
124 https://www.dfg.de/en/research_funding/programmes/nfdi/index.html
125 https://quest-dashboard.charite.de/#tabStart
126 https://medbib-charite.github.io/oa-dashboard/
127 https://doi.org/10.6094/UNIFR/245817
128 https://www.uni-heidelberg.de/en/university/about-the-university/good-academic-practice/open-access-policy
129 https://www.osc.uni-muenchen.de/index.html
130 https://www.osc.uni-muenchen.de/toolbox/index.html

Country	National policy	Monitoring of national policy	Monitoring of OS	LERU Institutions
Ireland	National action plan[131] and National Framework on the Transition to an Open Research Environment[132], developed by National Open Research Forum (NORF).	Currently being developed[133]	No	**Trinity College Dublin**, Policy:TA OA policy[134]
Italy	In 2022 the Italian Ministry of University and Research (MUR) published the National Plan for Open Science[135] (available in Italian and English[136])	No monitoring at the moment. A working group was appointed to establish guidelines for monitoring	Only at local level[137] (monitoring according to the 8 pillars of OS) No, but to be developed	**University of Milan**, Policy: RDM policy[138] OA policy[139]
Spain	A four-year National Open Science Strategy (ENCA)[140] (2023-2027) was approved in May 2023, including a budget of €23.8M per year. The ENCA strategic axes are: Digital infrastructures for open science, Management of research data following the FAIR principles, Open access to scientific publications and Incentives, recognition and training.	Currently only monitoring of OA levels in universities (by national libraries), proposal to build a national portal		**Universitat de Barcelona**, Policy: OA policy[141]

131 https://norf.ie/national-action-plan/
132 https://repository.dri.ie/catalog/0287dj04d
133 https://norf.ie/national-open-access-monitoring/
134 http://www.tara.tcd.ie/handle/2262/80574
135 https://www.researchitaly.mur.gov.it/en/2022/07/15/national-plan-for-open-science-published-by-the-mur/
136 https://www.mur.gov.it/sites/default/files/2023-01/PNSA_2021-27_ENG.pdf
137 https://unimibox.unimi.it/index.php/s/9WsbKCCEzm7trri
138 https://www.unimi.it/en/node/15884
139 https://www.unimi.it/en/node/12941
140 https://www.ciencia.gob.es/en/Noticias/2023/mayo/El-Gobierno-aprueba-la-primera-Estrategia-Nacional-de-Ciencia-Abierta.html
141 https://diposit.ub.edu/dspace/handle/2445/142065

Country	National policy	Monitoring of national policy	Monitoring of OS	LERU Institutions
Sweden	Swedish Government, the Swedish Research Council (mandate for Open Access to Research Data[142]), the National Library (mandate for Open Access to Scientific Publications[143]), and The Association of Swedish Higher Education Institutions	The National Library of Sweden (NLS) monitors the work for open access to scholarly publications and publishes a yearly report[144] (in Swedish) with an assessment of the current situation. NLS also monitors the total cost of publishing at Swedish universities on a yearly basis and publishes a separate report[145] on this (in Swedish).	The Association of Swedish Higher Education Institutions monitors how Swedish universities handle research data, and continuously evaluate and, if necessary, update the road map and action plan for OS[146]. The Swedish Research Council publishes a yearly report[147] (in Swedish, summary in English) which presents a combined mapping, analysis and assessment of the national work with open access to research data. There is, however, presently no reliable statistics describing the level of compliance with FAIR principles among Swedish universities.	**Lund University**, Policy: Research strategy for Lund University 2023-2026[148], OA policy[149]
Switzerland	- Swiss Open Science Policy is centred around Open Access[150] (Action Plan[151]; Implementation Plan[152]) and Open Research Data[153] (Action Plan[154]). - The Swiss Universities Open Science Programme for 2021-2024 includes funding to support policy implementation[155].	The Action Plan includes monitoring. This has not yet begun but will be overseen by the Swiss Universities' Open Science Delegation[156]	Not known	**Université de Genève**, Policy: OA policy[157] **University of Zürich**, Policy: OS policy[158]

142 https://www.vr.se/english/mandates/open-science/open-access-to-research-data.html
143 https://www.kb.se/samverkan-och-utveckling/oppen-tillgang-och-bibsamkonsortiet/oppen-tillgang.html
144 https://urn.kb.se/resolve?urn=urn:nbn:se:kb:publ-660
145 https://urn.kb.se/resolve?urn=urn:nbn:se:kb:publ-710
146 https://oppenvetenskap.se/
147 https://www.vr.se/download/18.72c4495e17f44b64443b03a/1647009787100/Samordningsuppdrag%20om%20%C3%B6ppen%20tillg%C3%A5ng%20till%20forskningsdata%20VR%202022.pdf
148 https://urn.kb.se/resolve?urn=urn:nbn:se:kb:publ-660
149 https://www.staff.lu.se/sites/staff.lu.se/files/2021-09/Open-access-policy-for-publications-and-artistic-works.pdf
150 https://www.swissuniversities.ch/themen/digitalisierung/open-access
151 https://www.swissuniversities.ch/fileadmin/swissuniversities/Dokumente/Hochschulpolitik/Open_Access/Plan_d_action-f.pdf
152 https://www.swissuniversities.ch/fileadmin/swissuniversities/Dokumente/Hochschulpolitik/Open_Science/PgB_OpenScience_-_Implementation_Phase_A_2021-2024_v6.4.pdf
153 https://www.swissuniversities.ch/themen/digitalisierung/open-research-data
154 https://www.swissuniversities.ch/fileadmin/swissuniversities/Dokumente/Hochschulpolitik/ORD/ActionPlanV1.0_December_2021_def.pdf
155 https://www.swissuniversities.ch/en/topics/digitalisation/open-science/programme
156 https://www.swissuniversities.ch/en/organisation/bodies/delegations/delegation-open-science
157 https://www.unige.ch/biblio/en/openaccess/understand/open-access-policy/
158 https://www.openscience.uzh.ch/en/definition/policy.html

Country	National policy	Monitoring of national policy	Monitoring of OS	LERU Institutions
The Netherlands	- Together, all public stakeholders in research have set an ambition for open science in 2030[159]. - National policy to provide 100% open access in 2020; Dutch national funders implemented Plan S; since 2019 national initiative to reform Recognition and Rewards led by the federation of universities (UNL); national government dedicates for the coming ten years €20M yearly to support the transition to open science; national effort to coordinate the transition to Open Science (Open Science NL[160])	- National monitoring only for OA for peer-reviewed journal articles[161] - National repositories that monitor data sets by license, openness, etc. An example is DANS (Dutch Archiving and Networking Services[162]).	Yes, partly via Strategy Evaluation Protocol[163] to be expected as one of the goals of Open Science NL.	**University of Amsterdam**, Policy: OA policy[164] **Leiden University**, Policy: OS program[165] OA policy[166] **Utrecht University**, Policy: OS program[167] OA policy[168]

159 https://doi.org/10.5281/zenodo.7433767
160 https://www.openscience.nl/
161 https://www.openaccess.nl/en/in-the-netherlands/monitor
162 https://dans.knaw.nl/en/
163 https://www.universiteitenvannederland.nl/onderwerpen/onderzoek/evaluatie-protocol-onderzoek-sep
164 https://uba.uva.nl/binaries/content/assets/subsites/bibliotheek/open-access/20210101_open_access_policy_uva_en.pdf
165 https://www.universiteitleiden.nl/en/dossiers/academia-in-motion/open-science
166 https://www.library.universiteitleiden.nl/researchers/open-access/background-and-policies
167 https://www.uu.nl/en/research/open-science
168 https://www.uu.nl/en/university-library/advice-support-to/researchers/publishing-support/open-access/open-access-policy-strategy

Country	National policy	Monitoring of national policy	Monitoring of OS	LERU Institutions
UK	- Major UK funders have OS policies that require all of their funded research to be published open access, including giving copyright to institutions, rather than journals, via a CC-BY license. This is true for UK Research and Innovation[169], The Wellcome Trust[170], the British Heart Foundation[171], and Cancer Research UK[172]. - The UK Reproducibility Network also has a programme to promote open research[173].	No official monitoring yet, but several relevant initiatives: - UKRI is working on a 'Monitoring and Evaluation Framework'[174], but so far none of the Research Councils monitor compliance. - UK Reproducibility Network includes evaluation[175] in its open research programme, including an Open Research Survey[176]. - Several initiatives in the area, e.g. via JISC[177] and in particular domains, e.g. via ELIXIR-UK[178]. - Monitoring of APC grants and principles of DORA (by UKRI and Wellcome) - The Fairsharing.org project[179] (based at the University of Oxford) is collating policy examples. - Scottish Government's RESAS[180] research programme adopted a new Data Management mandate for its 2022-2027 funding programme[181]	- Yes, via UK REF: requires publications to be made open access, and includes an assessment of compliance (for 2021)[182]	**Imperial College London**, Policy: RDM policy[183] OA policy[184] **University of Cambridge**, Policy: OS position statement[185] **University College London**, Policy: Statement on transparency in research[186] RDM policy[187] Bibliometrics policy[188] OA policy[189] **University of Edinburgh**, Policy: OS roadmap[190] RDM policy[191] Publications and copyright policy[192]

169 https://www.ukri.org/publications/ukri-open-access-policy/
170 https://wellcome.org/grant-funding/guidance/open-access-guidance/open-access-policy
171 https://www.bhf.org.uk/for-professionals/information-for-researchers/managing-your-grant/open-access-policy
172 https://www.cancerresearchuk.org/funding-for-researchers/applying-for-funding/policies-that-affect-your-grant/policy-on-open-access
173 https://www.ukrn.org/open-research-programme/
174 https://www.ukri.org/what-we-offer/supporting-healthy-research-and-innovation-culture/open-research/open-access-policies-review/implementing-our-open-access-policy/
175 https://www.ukrn.org/ws2-evaluation/
176 https://osf.io/preprints/metaarxiv/w48yh/
177 https://www.jiscmail.ac.uk/
178 https://elixiruknode.org/
179 https://fairsharing.org/
180 https://www.gov.scot/publications/environment-agriculture-and-food-strategic-research-main-research-providers/
181 https://www.gov.scot/publications/environment-agriculture-and-food-strategic-research-2022-27-overview/pages/strategic-research-programme-2022-to-2027/
182 https://results2021.ref.ac.uk
183 https://www.imperial.ac.uk/media/imperial-college/research-and-innovation/research-office/public/Imperial-College-RDM-Policy.pdf
184 https://www.imperial.ac.uk/research-and-innovation/support-for-staff/scholarly-communication/open-access/oa-policy/
185 https://osc.cam.ac.uk/open-research-position-statement
186 https://www.ucl.ac.uk/research/sites/research/files/ucl_statement_on_transparency_in_research_november_20191.pdf
187 https://www.ucl.ac.uk/isd/sites/isd/files/ucl_research_data_policy_v6.pdf
188 https://www.ucl.ac.uk/research/strategy-and-policy/bibliometrics-ucl
189 https://www.ucl.ac.uk/library/open-science-research-support/open-access/ucl-open-access-requirements
190 https://www.ed.ac.uk/sites/default/files/atoms/files/edinburgh_open_research_roadmap_jan2023_v1-1.pdf
191 https://www.ed.ac.uk/information-services/about/policies-and-regulations/research-data-policy
192 https://www.ed.ac.uk/information-services/about/policies-and-regulations/research-publications

Country	National policy	Monitoring of national policy	Monitoring of OS	LERU Institutions
				University of Oxford, Policy: RDM policy[193] OA policy[194]
Europe	- European Union Open Science Policy[195] - Horizon Europe Open Science Policy[196]	Monitored through EC Open Science Monitor?[197]	European Commission open science monitor[198]	European universities and EC grant recipients
USA	Research funded by National Institutes of Health (since 2008) must comply with its Public Access policy[199]: any publications resulting from NIH-funded research must be deposited in PubMed Central. This does not require a CC-BY license and thus is not exactly the same as Open Access	NIH Public Access Compliance Monitor[200]	NIH Public Access Compliance Monitor[201]	NIH Grant Recipients

193 https://researchdata.ox.ac.uk/university-oxford-data-management-policy
194 https://openaccess.ox.ac.uk/wp-content/uploads/sites/2/2022/11/University-of-Oxford-OA-Publications-Policy-2022-v1-1.pdf
195 https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science_en
196 https://op.europa.eu/en/web/eu-law-and-publications/publication-detail/-/publication/9570017e-cd82-11eb-ac72-01aa75ed71a1
197 https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science/open-science-monitor_en
198 https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science/open-science-monitor_en
199 https://publicaccess.nih.gov/
200 https://www.ncbi.nlm.nih.gov/pmc/utils/pacm/static/pacm-user-guide.pdf
201 https://www.ncbi.nlm.nih.gov/pmc/utils/pacm/static/pacm-user-guide.pdf

On an institutional level, Leiden University[202] and Utrecht University[203] are two examples of universities with clear and informative web resources providing examples of good practice related to different open science themes. Some universities place their trust in the commitment of their communities. For example, Sorbonne University's charter for open access to publications encourages the university to provide training for its community, equipping them to effectively address the challenges of open science.

On a national level, several countries have clear Open Science policies that include dedicated funding, including Finland[204], France[205], Spain[206], and Switzerland[207]. Finland is also notable for its well-developed, clear approach to monitoring of Open Science[208]. It should be noted that open science is not limited to Europe or the USA. For example, Latin America and the Caribbean have a strong tradition in open science, fair data and citizen science[209].

When it comes to access to regularly updated road maps that monitor the progress toward open science, good examples can be found at University of Edinburgh[210] and Utrecht University[211]. The University of Edinburgh's roadmap is also notable for adding 'Cultural Change' as a ninth "pillar of open science", underscoring the concept that cultural change is essential to realise the many benefits of open science.

At institutional, national, and international levels, it is evident that the most well-established open science policies and monitoring primarily concern Open Access to publications. This includes the maintenance of regularly updated dashboards for reporting the percentage of open access publications, as exemplified by Charité Berlin[212] and the Curtin Open Knowledge Initiative[213]. While some universities and countries have implemented policies related to FAIR Data, these are less prevalent than those for Open Access. It is worth noting that France has its own national open data warehouse. The French Open Science monitor also estimates the percentage of publications with Open data (using a text analysis tool from DataSeer)[214], and the Charité Berlin (Germany) also does this for open data and preregistration[215]. Formal monitoring for open data, however, is not yet widespread.

When it comes to the other six pillars of open science[216] (Education and Skills, Rewards and Incentives, Next-generation metrics, Research Integrity, Citizen Science, and The European Open Science Cloud), formal policies and monitoring have not always been established. From an Open Science perspective, FAIR data is a fundamental tool for enhancing research integrity and reproducibility.

Our LERU working group's objective is to identify next-generation metrics to evaluate and incentivise open science practices. It is important to note that many of the eight open science pillars intersect. For instance, progress in Open Access and FAIR data can contribute to enhancing Research Integrity. Therefore, it might not be necessary to create separate policies for each pillar, but rather consider how open science policies can collectively advance multiple pillars simultaneously.

9.4 Recommendations

Whether on an institutional, national, or international level, well-structured and updated web resources are key to communicate and promote information and services related to open science; the CoNOSC website, mentioned above, is one good example of this. Such web resources showcase the progress of open science within each country or institution and helps to identify best practice for those whose open science policies or programmes are less developed. Such information benefits users both within and beyond each institution. The recommendations derived from the survey results, as summarised in the table above, are as follows:

- All LERU members should have websites that describe the institutional and national policies for open science (including OA and FAIR data). These should be in multiple languages (including English) to ensure that they are accessible.
- All countries should have open science dashboards similar to that from Charité Berlin or the French Open Science Monitor, i.e., reporting not only OA publications, but also preprints, open data, preregistration, etc.

202 https://www.universiteitleiden.nl/en/dossiers/academia-in-motion/open-science
203 https://www.uu.nl/en/research/open-science/good-practices-and-illustrations
204 https://avointiede.fi/sites/default/files/2020-02/declaration2020_0.pdf
205 https://www.ouvrirlascience.fr/second-national-plan-for-open-science-2021-2024/
206 https://www.ciencia.gob.es/en/Noticias/2023/mayo/El-Gobierno-aprueba-la-primera-Estrategia-Nacional-de-Ciencia-Abierta.html
207 https://www.swissuniversities.ch/en/topics/digitalisation/open-science/programme
208 https://avointiede.fi/en/policies-materials/monitoring
209 https://www.unesco.org/en/articles/open-science-latin-america-and-caribbean-strong-tradition-long-journey-ahead
210 https://www.ed.ac.uk/information-services/research-support/open-research/about
211 https://www.uu.nl/en/research/open-science/knowledge-center/background-and-visiondocuments
212 https://medbib-charite.github.io/oa-dashboard/
213 https://open.coki.ac/
214 https://frenchopensciencemonitor.esr.gouv.fr/
215 https://quest-dashboard.charite.de/#tabStart
216 https://www.leru.org/publications/open-science-and-its-role-in-universities-a-roadmap-for-cultural-change

In conclusion, the creation, advancement, and implementation of next-generation metrics necessitate the support and acceptance of both the academic and policy communities. This is a multifaceted undertaking that demands an ongoing process of social acceptance. Such processes are well articulated in a recent book, 'Gradual' by Greg Berman and Aubrey Fox, published by Oxford University Press in 2023. In the book, the authors detail how decisions involving the social acceptance of intricate policy instruments evolve through incremental implementation, which is reflected in the book's title, 'Gradual.' The development and implementation of next-generation metrics may encounter a similar reality.

10. The challenge of data availability for next generation metrics

10.1 General aspects of data needs and availability

As institutions introduce new values, goals, and policies, there is an increasing demand for data to inform, assess, and monitor their progress. Given the diversity of values, goals, and policies across research teams, departments, institutions, and countries, it becomes evident that a one-size-fits-all, non-customizable metric cannot adequately address the current demand. Positive developments are emerging on the supply side. We are witnessing an increase in aggregators that offer machine-readable knowledge graphs with granular data, aiming to comprehensively capture research activities and outputs. Examples include OpenAlex, OpenAire, The LENS, Dimensions, and more narrowly focused ones such as COKI and DataCite Commons. Some of these are fully free, with open-licensed data, whereas some are more closed. More generally, we also see improvements, such as the uptake of persistent identifiers, broader coverage of databases, easier machine access to data through APIs, or community governance of services. It is important to note that, as we write, many of these initiatives and improvements are under construction, ongoing, and not yet implemented across the board. There are many dependencies, and it is not possible for one organisation on its own to provide a service that does it all. The knowledge graphs mentioned primarily adopt a strategy of gathering extensive data on outputs from numerous organisations, which are often covered by various registries. They then employ various technologies to clean, deduplicate, connect, and enrich the data. However, this approach results in these knowledge graphs, despite their larger size, having a greater diversity of data quality and more varying degrees of completeness in their records compared to traditional, narrower, and more selective databases. Additionally, they require more comprehensive documentation to make sense of the data and address biases and data gaps. Further improvements also rely on the adoption of standards by all stakeholders involved in research, including funding organisations and publishers.

It is essential to distinguish between the availability of (i) granular data (primarily characteristics of researchers, organisations, and their research activities and outputs), (ii) metrics (calculations and combinations of data for specific use cases, enabling comparisons over time, against a baseline, or among individuals, organisations, or other groupings), and (iii) code and software used to generate or analyse data. In this discussion, our primary focus is on the availability of granular data.

This approach, centred on data rather than specific metrics, aligns well with current needs, enabling organisations to conduct evaluations that are purpose-specific and closely aligned with their values and goals. When data availability and standards are high, and data is freely reusable, conducting assessments becomes a matter of creating (or reusing and adapting) code for data analysis. However, this process is not without its complexities and requires the expertise of software engineers within research intelligence communities. It may also necessitate, to some extent, a cultural shift within these communities, as the emphasis shifts from primarily supporting and executing summative assessments to embracing more formative assessments and customised approaches.

Such an approach, using open disaggregated data and custom code for analysis to inform formative assessments, will require several years for full development. It may be crucial to seek practical and user-friendly solutions based on stable shared code and well-documented methods, striking a balance between simplicity and the absence of impact factors, H-indexes, and rankings on one hand, and the extensive pools of disaggregated data that need interpretation on the other. Institutions, including LERU institutions but also many other institutions, could collaborate to create code and methods applicable to similar situations and recurring needs. Communities of research information professionals, such as Eurocris, could also play a role in this development.

While it is valuable to have a general understanding of how data availability is evolving, this broad overview may lack meaningful insights when applied to specific use cases. For instance, institutions may seek to gain insights into the extent and impact of data sharing in their publicly funded projects, which requires a more tailored approach. Due to the scope limitations of this report, we cannot delve into all possible use cases, each with its unique combination of data types (as indicated in the top yellow block in Figure 3) and specific additional requirements. Instead, we take a step back to explore the broader spectrum of new metrics we might wish to measure, particularly in the context of open science. We address additional requirements separately.

10.2 Quantitative insights into research-related data availability

The set of new metrics we may wish to measure encompasses various types of outputs, including non-English content, collaboration data, and impact indicators. Among these, output types are the most straightforward as they represent tangible entities, such as files published or shared with URLs, DOIs, or other identifiers. With current trends, data related to these outputs is likely to improve further over time.

However, certain challenges will persist. Output that is not digital, not available online, lacks identifiers, or is shared on less professional platforms or by less reputable publishers may remain 'in the dark.' Aggregating data on these outputs in an automated manner can be challenging, especially when the information is absent in institutional CRIS systems or personal ORCID profiles. For any meaningful assessment initiative, it is crucial to explicitly address this 'long tail' of more elusive outputs and activities, rather than simply ignoring them.

The availability of collaboration data relies on the presence and quality of affiliation and researcher/author data in records of grants, projects, and research outputs, ideally based on ROR and ORCID identifiers. A collaboration can be identified when any grant, project, or research output is linked to more than one institution (ROR) or researcher/author (ORCID). While collaborative groups between individuals and institutions, such as teams, departments, labs, or project groups, may be identifiable in CRIS systems, they are often not available in public data. These groupings typically lack a specific group identifier, with the only indirect identification being through a project or grant identifier. Conducting formative assessments at this intermediate level can be highly valuable, but these groupings tend to be dynamic and can change over time. As a result, assessments will often need to be customised to the specific group composition and goals, relying on the availability of granular data that includes person identifiers, or on data from the institution's CRIS system if it can identify specific groups.

Another general notion on data availability is that in various systems good data on disciplines is lacking. This is problematic because research cultures and practices differ between the various disciplines and for policies to be acceptable and effective and open science practices gain traction, disciplinary approaches and discussions in disciplinary communities are essential. For that we also need disciplinary insights. Whether manually curated, automatically assigned based on factors like abstracts, or assigned by proxy based on sharing platforms or venues. However, apart from the generation and availability of data, the choice of what and how to assess also varies depending on the discipline. While STEM culture and assessment procedures are dominant, initiatives such as HuMetricsHSS[217] show that disciplinary communities can and should have a voice in how data and metrics inform assessment.

There is growing awareness of the data-related aspects of enhancing research assessment. Ongoing efforts are being made in various contexts to improve these situations. The European projects OPUS[218], GraspOS[219] and PATHOS[220] are actively addressing the information needed to comprehend the process and outcomes of open science.

The OPUS project is focused on reforming research and researcher assessment at Research Performing Organisations (RPOs) and Research Funding Organisations (RFOs) to promote Open Science practices. As part of its research assessment framework (RAF) it has developed an extensive list[221] of indicators, covering research, education, leadership, and valorisation, which can serve as a valuable resource for selecting relevant assessment criteria. The GraspOS project aims to explore responsible research assessment methods that consider Open Science practices. It is also building a federated infrastructure, functioning as an open data space, providing data, indicators, tools, services, and guidance to support policy reform in research assessment.

In contrast to the OPUS Project, GraspOS is dedicated to building and operating a data infrastructure, with a focus on extending the European Open Science Cloud and utilising the OpenAire data infrastructure. PATHOS has a specific objective of gathering evidence on open science and its academic, societal, economic impacts, as well as reproducibility. It has produced the Open Science Indicator Handbook[222], which provides information on available data. These projects collectively showcase the enthusiasm and potential for developing new metrics and assessment methods to promote open science.

In addition to the three European projects, there are other developments in the research information space that contribute to defining, selecting, and testing indicators, along with improving technical aspects to enhance efficiency. These efforts are carried out within various national or institutional open science monitors, FAIRness assessments for software and data, and through PID-graphs, such as those offered by OpenAlex, DataCite, and OpenAire. These PID-graphs connect data using persistent identifiers for authors, objects, organizations, projects, grants, funders, containers such as journals, and so on.

217 HuMetricsHSS. Humane Metrics Initiative. https://humetricshss.org/
218 OPUS, https://opusproject.eu/
219 GraspOS, https://graspos.eu/
220 PATHOS, https://pathos-project.eu/
221 OPUS deliverable 3.1: Indicators and metrics to test in the pilots, https://opusproject.eu/wp-content/uploads/2023/09/OPUS_D3.1_IndicatorsMetrics_FINAL_PUBLIC.pdf
222 PATHOS Open Science Indicator Handbook, https://handbook.pathos-project.eu/

Finally, the Delphi Survey[223] aimed to establish a consensus on the core open science practices for biomedical research. While it may not fully address the needs of other academic disciplines, it is notable for its broad geographical representation, incorporating input from institutions across Africa, Asia, Australia, Europe, North America, and South America. The authors of the Delphi Survey are currently developing a digital dashboard to track progress on these core practices across numerous international research institutions.

Table 2 Meeting New Data Needs - Progress and Ongoing Challenges

What we want to measure	What is available or technically possible, what is improving	What is not yet (fully) available or not yet (fully) solved
open access status of outputs	- OA-detection (e.g. UPW, including OA-types) - open reuse license detection - output PIDs (DOI, URN, etc.) - data/software/preprint citations (bi-directional)	- diamond OA detection - OA detection for non-journal and non-DOI output
data FAIRness and availability/reuse	- more inclusive databases (doctype/lang./geogr.) - open reuse license detection - output PIDs (DOI, URN, etc.) - data/software/preprint citations (bi-directional)	- data on FAIRness - data on citation/reuse
code & software availability/reuse	- more inclusive databases (doctype/lang./geogr.) - open reuse license detection - output PIDs (DOI, URN, etc.) - data/software/preprint citations (bi-directional)	- archiving of software/code - ORCID in GitHub
preprinting	- more inclusive databases (doctype/lang./geogr.) - open reuse license detection - output PIDs (DOI, URN, etc.) - data/software/preprint citations (bi-directional)	- bidirectional links between preprints and formal publications/reviews of those
conference output	- more inclusive databases (doctype/lang./geogr.) - open reuse license detection - output PIDs (DOI, URN, etc.)	- DOIs for conference material, at level of papers
books, chapters	- more inclusive databases (doctype/lang./geogr.) - open reuse license detection - output PIDs (DOI, URN, etc.)	- DOI for books (at chapter level for edited volumes); citations for books
output in all languages	- more inclusive databases (doctype/lang./geogr.)	- inclusion of non-English output in scholarly databases - funding/expertise for PID registering by (smaller) publishers and platforms
(societal) impact	- open reuse license detection - data of research/output mentions beyond academia - funding/funder PIDs (funder registry, grant IDs) - data/software/preprint citations (bi-directional) - AI-based analyses/data	- altmetrics coverage - understanding meaning of mentions
collaborations	- more inclusive databases (doctype/lang./geogr.) - (sub)organization PIDs (ROR) - creator PIDs (ORCID) - project PIDs (RAID) - role indicators (e.g. using CREDIT taxonomy)	- uptake of project PIDs - development of suborganisation RORs - better detection coverage of funding/grant information

223 Delphi Survey, https://doi.org/10.1371/journal.pbio.3001949

Despite the ongoing projects and improvements, more work is required to enhance data availability. **Table 2** details the new data needs that have been addressed by improvements in availability and highlights the areas that are still lacking or unresolved. The key areas still in need of expansion in our measurement efforts include greater geographical and disciplinary inclusivity in databases, increased adoption of persistent identifiers (PIDs), particularly for researchers (ORCID) and projects (RAID), and improved, open altmetrics data that can offer evidence of research mentions or usage.

10.3 Measuring things differently: opportunities, requirements, and challenges of next-generation metrics data sources

Measuring different things primarily involves data supply, whereas measuring things differently is about ensuring that the quantification of the research process, outputs, and outcomes is carried out responsibly and aligns with the goals of research communities (see **Table 3**). To conduct meaningful assessments that contribute to research improvement, we require insights at the project and team levels, which naturally involve the research teams under assessment.

There is a growing awareness of the significance of transparency, openness, and academic governance of essential data services. Some of these data sources are already publicly controlled, often by the research community, such as ROR, ORCID, DataCite, and OpenAire Others, like OpenAlex, are commercial but adhere to open principles. Many have conducted public self-assessments[224] of their alignment with the Principles of Open Scholarly Infrastructure[225]. However, there are numerous commercial providers that still play a central role in research assessment, including Elsevier with Scival, Scopus, and PlumX, Clarivate with Web of Science, InCites, and Publons, and Digital Science with Dimensions and Altmetrics. To enable customised and formative assessments, often at new units of analysis, the handling of disaggregated open data is crucial. Indeed, one of the objectives of the GraspOS project (Next Generation Research Assessment to Promote Open Science) is to establish an open metrics infrastructure that includes open data sources for metrics. Responsible assessments require full transparency in all cases, allowing the combination of data from various sources without being constrained by a fixed set of metrics and closed algorithms. It's imperative that (meta)data from any data source used be fully open to verify its relevance, quality, and completeness.

Combining qualitative and quantitative data in a way that provides a deeper and more contextualised understanding of the scientific activity of the analysed units is a challenging yet crucial aspect of measuring things differently.

This approach serves as a tool for formative assessment and self-assessment, aligning with the unit's own objectives, as emphasised in various research evaluation and publication metrics recommendations.

Qualitative insights extend beyond traditional methods like peer review, assessment panels, interviews, or expert committees. They delve deeper into the nature and contexts of the research being assessed. Basic output metadata is insufficient for this purpose. Analysing abstracts, summaries, and full texts of outputs provides opportunities to assess research objectives and results in a more meaningful way, focusing on the nature of the research and its impacts rather than just quantities. However, this requires open availability of at least abstracts, and ideally full texts, in machine-readable formats. This approach necessitates further experimentation, with new technologies, including AI-based tools. These tools must undergo rigorous assessment before widespread use, ensuring they align with the basic SCOPE principles and avoiding a hasty technological fix.

To align with a more meaningful, qualitative, and formative assessment, there is a need to avoid rankings and composite metrics. Selecting a suitable set of indicators that provide information about various aspects of the analysed units yields a more comprehensive perspective than relying on a single indicator. Rather than ranking against competitors, assessment can shift its focus towards comparisons with internal goals, past achievements, or the expectations of societal stakeholders.

10.4 Recommendations

To facilitate measuring different aspects and adopting new measurement approaches, all stakeholders in the research information ecosystem have unique roles to play, and many have already been actively contributing to this evolving landscape. There are five categories of stakeholders, each with specific responsibilities and potential contributions to advancing research assessment through data.

1. **Platforms** (e.g., publication platforms, repositories): These entities should prioritise inclusivity, openness, machine-readable formats, and the use of persistent identifiers when hosting research content.
2. **Registries** (e.g., Crossref, DataCite, ROR, ORCID): Registries are vital for providing fast, comprehensive, and interoperable metadata, enabling efficient data sharing and discovery.
3. **Aggregators** (e.g., OpenAire, OpenAlex, CORE): Aggregators play a critical role in consolidating and harmonising research data, ensuring that the data remains open and free from proprietary metrics or restrictions.

224 POSSE, https://openscholarlyinfrastructure.org/posse/
225 Principles of Open Scholarly Infrastructure (POSI), https://openscholarlyinfrastructure.org/

4. **Services**: Organisations offering services should provide valuable insights into research patterns and processes, assisting in on-demand data analysis and interpretation.
5. **Research Assessment Experts**: These experts, working within research institutions alongside researchers, are responsible for defining assessment goals and methods. They should be prepared to adapt their practices to accommodate the changing landscape of research assessment. This may involve training and sharing best practices to ensure effective measurement of diverse research aspects and the adoption of alternative assessment approaches.

Table 3. Measuring approaches: current status and ongoing developments

How we want to measure	What is currently available or technically feasible / what is improving	What is not (yet) available or remains unsolved
data at team level	- creator PIDs (ORCID) - role indicators (e.g. using CREDIT taxonomy)	-
data at project level	- creator PIDs (ORCID) - project PIDs (RAID)	- uptake and support of RAID
community control & ownership of tools	- community governed tools/data	- number of infrastructures that have done self assessment on POSI principles
data supporting formative assessment	- more inclusive databases (doctype/lang./geogr.) - data of research/output mentions beyond academia - project PIDs (RAID) - role indicators (e.g. using CREDIT taxonomy) - AI-based analyses/data	- data providing early insights of usage - open usage data; - granularity of data
avoid ranking of data	- tools providing open granular data (API/dumps)	-
avoid composite metrics	- tools providing open granular data (API/dumps)	-
fully open and transparent data	- open citations (COCI) and open abstracts (I4OA) - open-source tools (e.g. OpenAlex)	- openness of abstracts - openness of altmetrics data
more qualitative data/insights (i.e. data about the nature of the research and nature of impacts)	- open citations (COCI) and open abstracts (I4OA) - open-source tools (e.g. OpenAlex) - AI-based analyses/data - text mining techniques for full text analyses	- openness of abstracts - openness of altmetrics data

Figure 3. Changes in Demand and supply research evaluation in the context of open science and new R&R

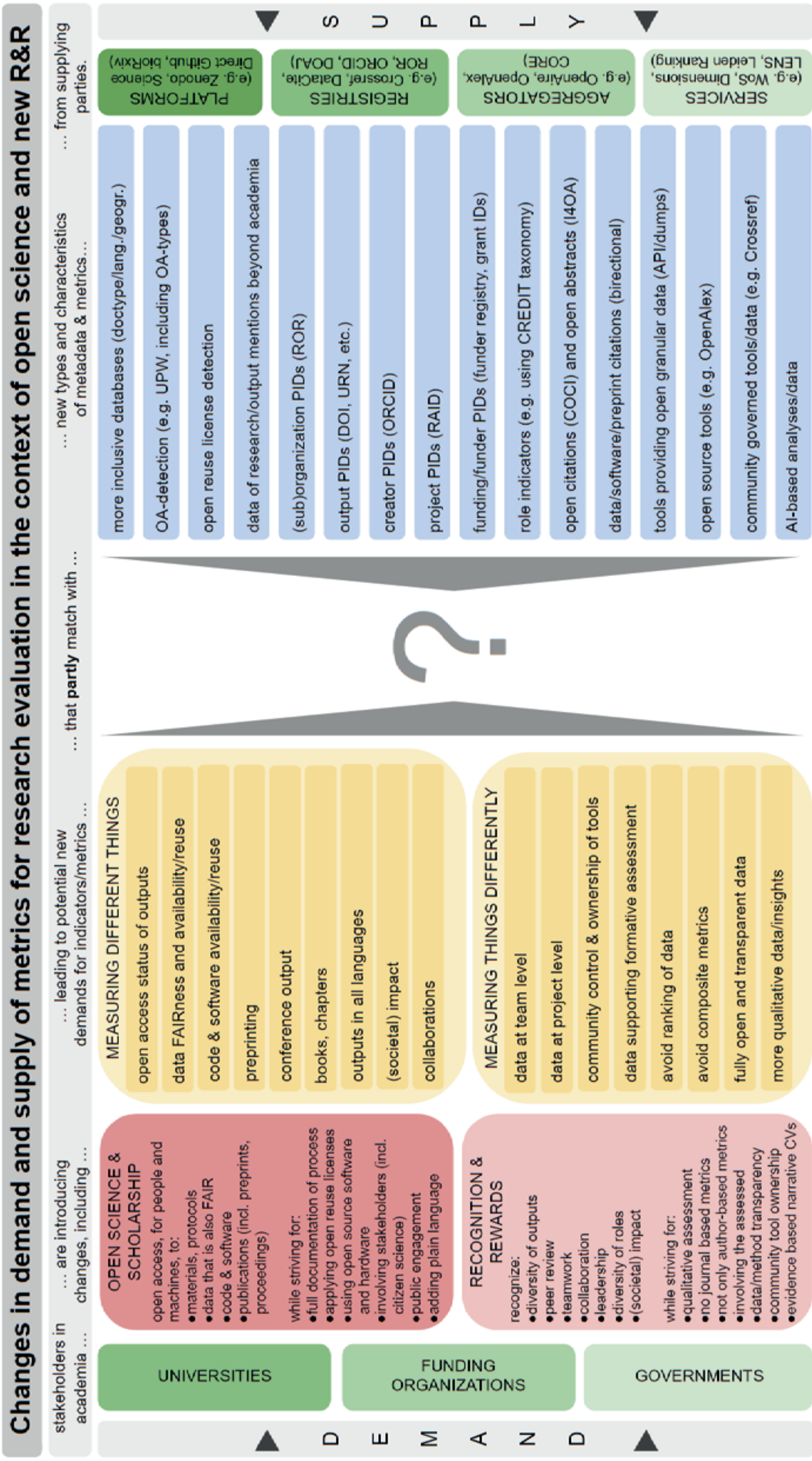
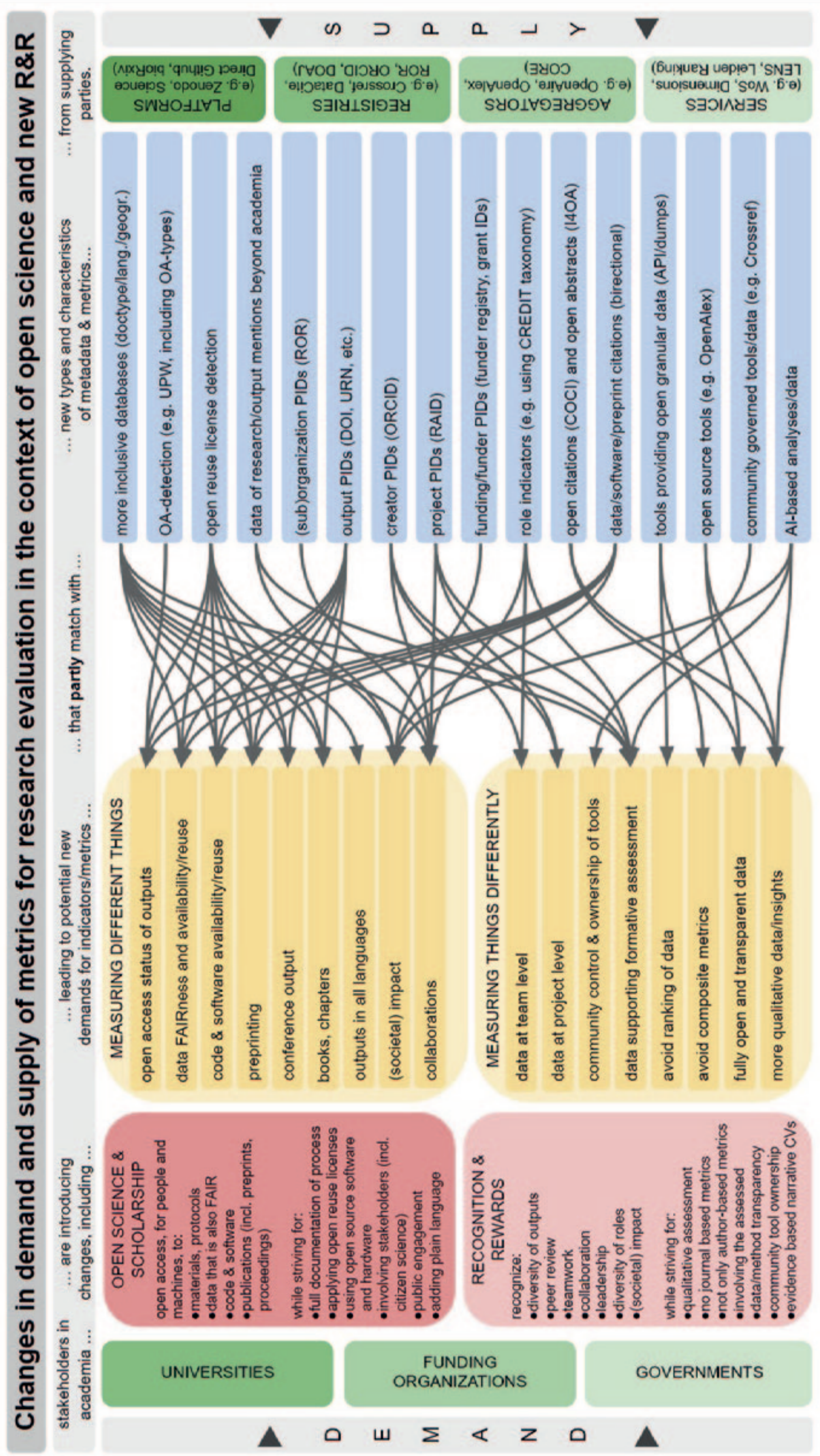


Figure 3. Changes in Demand and supply research evaluation in the context of open science and new R&R



Appendix I: Metrics terminology

Term	Definition / Scope (Our Interpretation)
Indicators (term used by e.g. Hong Kong principles / Leiden Manifesto)	Any clearly defined and collected attribute that can be used to compare the state or level of something relevant to the goals of an organisation or person. The aim is to use the insights gained either for summative purposes, such as selection and assessment, or for formative purposes, as input for development.
Metrics (term used by e.g. DORA / Altmetrics manifesto)	Indicators that are quantitative and mostly numeric, sometimes related to absolute or relative quantitative goals (e.g., 'we aim to increase x,' 'want to reach 80% by 2025,' etc.).
Research Metrics (term used by e.g. Leiden Manifesto)	Metrics informing studies of research, including but not limited to bibliometrics, encompass various aspects, such as funding and acknowledgements, and increasingly extend to a broader range of research practices, including data, code, and workflows. An important extension involves innovation metrics, such as technology transfer to industry, spin-offs, patents, and more.
Bibliometrics (term used by e.g. LERU)	Research metrics informing studies of research by measuring creation/authorship, output and usage of scholarly publications
Traditional Research Impact Metrics	Bibliometrics used to assess research impact, primarily through journal-level and article/book-level citation analysis
Technometrics	Technometrics is the systematic analysis and assessment of technological progress and innovation through the examination of patent data.
Databases	Databases are structured and organised collections of data that are generated as a direct outcome of research activities. These databases encompass various types of information, notably clinical research data, and extend to encompass specialised repositories such as tissue banks.
Next Generation Metrics (term used by e.g. EU, LERU,CESAER)	Next-generation metrics encompass both existing and novel indicators across various categories and sources. They are considered suitable for informing new research policy objectives and adhere to the latest insights and guidelines established over the past decade. These metrics are characterised by their relevance, validity, and responsible application in assessing research activities at different levels, aligning with contemporary research assessment frameworks and guidelines.
Alternative Metrics / Altmetrics (term used by e.g. LERU, EU, Altmetrics Manifesto)	Alternative research metrics offer an alternative or complement to traditional research impact metrics. They capture mentions and usage beyond mere citations in scholarly publications, providing a wider, faster, and more diverse set of impact indicators. While some altmetrics can be relevant for next-generation metrics, there are concerns about the reliability and validity of altmetrics indicators. Therefore, next-generation metrics should address these issues and develop a set of indicators and metrics that meet the expectations of reliability and validity.
New Metrics	New metrics is a broad term that encompasses various approaches and indicators used to assess research and scholarly impact. It is not synonymous with altmetrics, although altmetrics can be considered one category of new metrics. New metrics include both traditional and non-traditional indicators, and they go beyond the traditional citation-based metrics like the Journal Impact Factor.
Responsible Metrics (term used by e.g. Universities UK)	Research metrics that meet requirements for robustness, humility, transparency, diversity, and reflexivity, based on the principles outlined in 'The Metrics Tide'.
Webometrics	Metrics that track the visibility or references of academic activities and outputs on the web, as opposed to within scholarly literature. Altmetrics can be considered a subset of webometrics.

Appendix II: Recommendations from previous declarations and reports.

This report comprises sections that delve into the history of scientometrics, their appropriate and inappropriate applications in assessing individuals, teams, and institutions, and the prospects and challenges associated with next-generation metrics. Noteworthy previous recommendations and reports aimed at addressing these issues are outlined below.

In 2018, LERU formulated the following recommendations concerning the responsible use of metrics and indicators in its **LERU Open Science Roadmap**[226]:

- Develop a bibliometrics policy rooted in the principles of the Leiden Manifesto, aiming to foster a cultural shift within the academic community regarding research assessment.
- Integrate new forms of research evaluation into internal processes related to promotions, rewards, and research assessment.
- Create guidelines for research administrators and academics on best practices and pitfalls in traditional bibliometrics and the development of new metrics, in collaboration with the scientific community.
- Provide training, particularly for junior researchers and early-stage doctoral researchers, to facilitate their embrace of the cultural and practice changes associated with the responsible use of metrics (aligning with recommendations on education and skills).

These recommendations stem from a prolonged debate that commenced in the early 2010s concerning the necessity of innovative approaches in research evaluation and the professional advancement of researchers. This discourse has grown in significance since its inception.

A significant recent milestone in this progression is the **Global Research Council Statement of Principles on Recognising and Rewarding Researchers**[227] on the 31st of May 2023, which introduced new principles for research assessment by science funders:

- Research assessment should adopt a comprehensive and inclusive approach, recognising the wide array of research activities and outputs.
- Researchers should undergo evaluation considering the entirety of their engagements.
- Assessment should primarily rely on a qualitative methodology, supplemented by the responsible and open utilisation of quantitative indicators.
- Existing assessment criteria and procedures should undergo regular re-evaluation and adaptation.
- The promotion of equity, diversity, and inclusion should guide responsible research approaches and practices in all research aspects.
- Enhancing researcher mobility and improving the compatibility of research systems must be a collective and global endeavour.

In 2012, the **San Francisco Declaration on Research Assessment (DORA)**[228] highlighted the urgent need to enhance the evaluation of scientific research output by funding agencies, academic institutions, and other stakeholders. DORA was initially created to combat the disproportionate reliance on the Journal Impact Factor (JIF), a metric at the journal level, when inappropriately used to appraise the quality of individual research articles. DORA outlines a comprehensive framework for action, encompassing four key objectives:

- Gain a comprehensive understanding of the obstacles that impede change.
- Engage in experimentation with various ideas and approaches across all organisational levels.
- Foster the development of a shared vision for research assessment during policy review and revision.
- Disseminate and communicate this vision within the academic community and beyond to other research institutions.

DORA has assembled a repository of innovative research assessment practices that institutions can utilise as a valuable resource.

A crucial recommendation from the **EU-funded consortium ACUMEN**[229] in 2014 was the adoption of a narrative CV, complemented by relevant metrics. This consortium was responsible for pioneering an initial framework aimed at the systematic creation of these narrative CVs.

226 Ayris, P., Lopez de San Román, A., Maes, K., & Labastida, I. (2018). Open science and its role in universities: A roadmap for cultural change. Leuven: LERU Office. https://www.leru.org/publications/open-science-and-its-role-in-universities-a-roadmap-for-cultural-change
227 Global Research Council (2023). Statement of Principles on Recognising and Rewarding Researchers. https://globalresearchcouncil.org/fileadmin//documents/GRC_Publications/SoP_Recognising_and_Rewarding_Researchers.pdf
228 American Society for Cell Biology. (2012). San Francisco declaration on research assessment (DORA). https://sfdora.org/read/
229 Wouters, Paul, Judit Bar-Ilan, Mike Thelwall, Isidro F. Aguillo, Ülle Must, Frank Havemann, Hildrun Kretschmer, et al. “Acumen Final Report.” Brussels: European Commission, 2014. http://cordis.europa.eu/result/rcn/157423_en.html

In 2015, the **Leiden Manifesto**[230] introduced a series of recommendations primarily directed towards the evaluation of individual scientists and research groups:

- Utilise quantitative evaluation as a complementary tool to support qualitative, expert assessment.
- Evaluate performance in alignment with the research missions of the institution, group, or researcher.
- Emphasise open, transparent, and straightforward data collection and analytical processes.
- Ensure that those being evaluated have the ability to verify the data and analyses.
- Consider variations by field in publication and citation practices.
- Base assessments of individual researchers on a qualitative assessment of their entire portfolio.
- Avoid undue precision and false accuracy in the assessment process.
- Recognise the systemic impacts of evaluations and indicators on the research ecosystem.

The **Metric Tide Report**[231], an initiative of the UK HEFCE, also issued a set of more general recommendations in the same year:

- The research community should cultivate a more refined and nuanced understanding of the capabilities and limitations of quantitative indicators.
- At the institutional level, leaders of Higher Education Institutions (HEIs) should develop a clear statement of principles outlining their approach to research management and assessment, encompassing the role of quantitative indicators.
- Within their institutions, research managers and administrators should advocate for these principles and promote the responsible use of metrics.
- Data providers, analysts, and creators of university rankings and league tables should work towards enhancing transparency and interoperability between different measurement systems.

A comprehensive policy framework that holds particular relevance for next-generation metrics was formulated by the International Network of Research Management Societies (INORMS) Research Evaluation Group (REG) in 2021. This group unites representatives from various research management societies worldwide, working towards achieving improved, equitable, and more meaningful research evaluation practices. The **SCOPE Framework**[232], crafted by the REG, serves as a practical approach for implementing responsible research evaluation principles, ensuring the design of robust assessments.

The SCOPE framework is built upon three core principles:

1. Evaluate only when essential. Evaluation may not always be the most suitable approach. In certain cases, it might be more effective to facilitate desired behaviours rather than evaluating them.
2. Collaborate with the evaluated parties. Any evaluation should be developed and interpreted collaboratively with the communities under assessment.
3. Harness evaluation expertise. We should apply the same level of rigor to our evaluations as we do in our academic research.

SCOPE is an acronym that represents a five-stage process as follows:

START with what you value: Begin by identifying and prioritising the values and objectives pertinent to your evaluation.

- Clearly articulate what you value about the entity being evaluated.
- Do not begin with what others value (external drivers).
- Avoid starting with available data sources (the 'Streetlight Effect').

CONTEXT considerations: Consider the specific contextual factors and conditions that may influence the evaluation.

- Ensure your evaluation is context specific.
- Determine WHO you are evaluating (consider entity size and discipline).
- Define WHY you are conducting the evaluation.

OPTIONS for evaluating: Investigate and select the most appropriate methods and indicators for conducting the evaluation.

- Explore both quantitative and qualitative evaluation methods.
- Exercise caution when attempting to represent qualities using quantities.

PROBE deeply: Delve into the evaluation process thoroughly, ensuring a comprehensive examination of the chosen criteria.

- Examine WHO your evaluation approach might unintentionally discriminate against.
- Investigate HOW your evaluation approach could be manipulated.
- Analyse the potential unintended consequences.
- Weigh the cost-benefit of the evaluation.

230 Leiden Manifesto for Research Metrics. 2015. http://www.leidenmanifesto.org/
231 Wilsdon, J., et al. (2015). The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management. https://doi.org/10.13140/RG.2.1.4929.1363
232 SCOPE Framework for Research Evaluation. https://inorms.net/scope-framework-for-research-evaluation/

EVALUATE your evaluation: After completing the evaluation, assess the process itself to identify areas for improvement or refinement.

- Assess whether your evaluation achieved its goals.
- Consider if it provided formative insights in addition to summative results.
- Utilise the SCOPE framework for evaluating your evaluation.

When employing quantitative methods, the SCOPE framework underscores the significance of adhering to responsible metrics principles. Specifically, it highlights that quantitative approaches, such as bibliometrics, become less suitable as the entity or sample size decreases. If there is no alternative but to use data for small sample sizes or entities, it is essential to share this data with the entities being evaluated for verification. Moreover, the evaluated entities should have the chance to offer free-text comments. A useful checklist when designing indicators is as follows: Indicators should be:

- VALID - reflecting the concept measured.
- UNDERSTANDABLE
- TRANSPARENT - data underlying criteria should be released, with clearly explained limitations and degrees of uncertainty.
- FAIR - systematic bias should be avoided.
- ADAPTIVE - updated when bias, abuse or other weaknesses become apparent.
- REPRODUCIBLE - those who use the indicator should be able to reproduce it.

Recommendations for Responsible Researcher Assessment and Metrics Use

In 2020, the **Hong Kong Principles for Researcher Assessment**[233] were published with a focus on research integrity:

- **Principle 1**: Assess researchers on responsible practices from conception to delivery, including the development of the research idea, research design, methodology, execution, and effective dissemination.
- **Principle 2**: Value the accurate and transparent reporting of all research, regardless of the results.
- **Principle 3**: Value the practices of open science (open research)—such as open methods, materials, and data.
- **Principle 4**: Value a broad range of research and scholarship, such as replication, innovation, translation, synthesis, and meta-research.
- **Principle 5**: Value a range of other contributions to responsible research and scholarly activity, such as peer review for grants and publications, mentoring, outreach, and knowledge exchange.

In summary, recommendations for the responsible use of research metrics have been developed, addressing concerns about misuse, individual researcher assessment, open science promotion, and research integrity. These recommendations centre around three key issues regarding both the *governance process* in which next-generation metrics are applied as well as the *selection and application of specific metrics and indicators*:

- consider the goal of the evaluation in which the specific metrics will be used.
- take into account the fit between the metrics properties and the level of aggregation (e.g. individual, group, institute, nation).
- consider both the context from which the metrics were extracted and the context in which they will be used.

These considerations are also essential in the development and application of next-generation metrics.

233 Moher D, Bouter L, Kleinert S, Glasziou P, Sham MH, Barbour V, et al. (2020) The Hong Kong Principles for assessing researchers: Fostering research integrity. PLoS Biol 18(7): e3000737. https://doi.org/10.1371/journal.pbio.3000737

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/publications.

LERU, Brussels, April 2024
Available at: ttps://doi.org/10.5281/zenodo.10569960

Graphic design: Altera
Printed 100% climate-neutral by Van der Poorten

PUSHING
THE FRONTIERS
OF INNOVATIVE
RESEARCH

LERU Office

Minderbroedersstraat 8
3000 Leuven
Belgium

info@leru.org
www.leru.org

@LERUnews