

## Response to HEFCE's call for evidence: independent review of the role of metrics in research assessment



30 June 2014

**Yes, Elsevier would be interested in participating in a workshop / event to discuss the use of metrics in research assessment and management.**

This response covers the following sections:

- A. Why is Elsevier responding?
- B. For what purposes are research metrics used?
- C. Guiding principles for use of metrics in research assessment
- D. A model for generating and using metrics in research assessment
- E. Implications for using metrics in research assessment
- F. Response to the specific questions posed by HEFCE

## A. Why is Elsevier responding?

Elsevier is the world-leading provider of scientific, technical and medical information products and services<sup>1</sup>, and is also a major investor in digital technology<sup>2</sup>. We have invested significantly in building **expertise to deliver tools and services that extract, analyse and visualise information about research trends and performance from our databases and other data sources. Research metrics permeate these tools and services.** Elsevier brings these credentials to bear in sharing our perspective on the role of metrics in research assessment.

The Elsevier Research Intelligence portfolio<sup>3</sup> embodies these tools and services, comprising database, benchmarking and research information management tools, and customised reports and analytical services. Our Research Intelligence portfolio is widely used because it is developed in close collaboration with the research community, and **our extensive experience working with the research community in the UK and beyond informs our response to this consultation.** In the UK, for instance, we:

1. Were the **exclusive provider of publication and citation data for the REF 2014**, and supported more than 80 universities. 20 universities submitted to the REF using Pure<sup>4</sup>, our research information management system.
2. Produced the report on the **International Comparative Performance of the UK Research Base** for the Department of Business, Innovation and Skills in 2013<sup>5</sup> and 2011.
3. Collaborate with 8 research-intensive universities on the **Snowball Metrics**<sup>6</sup> programme, which aims to develop the international standard for metrics used to build and monitor institutional strategies.

Our customers include:

1. **Policy-makers:** Department of Business, Innovation and Skills (BIS)<sup>5</sup>; British Council of Hong Kong; Department of Industry, Innovation, Science, Research and Tertiary Education (DIISRTE, Australia); Illinois Science & Technology Coalition (ISTC); the World Bank<sup>7</sup>; United Nations Educational, Scientific and Cultural Organization (UNESCO)<sup>8</sup>; Organisation for Economic Co-operation and Development (OECD)<sup>9</sup>.
2. **Funding bodies:** Higher Education Funding Councils for England (HEFCE), and for Wales (HEFCW)<sup>10</sup>; Research Councils UK (RCUK); Australian Research Council (ARC); Science Europe<sup>11</sup>; Fundação para a Ciência e a Tecnologia (FCT, Portugal); Abilitazione Scientifica Nazionale (ASN, Italy); Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Brazil); Japan Science and Technology Agency (JST); Department of Energy (US).

---

<sup>1</sup> A complete list of Elsevier's products and services can be viewed at <http://www.elsevier.com/elsevier-products>.

<sup>2</sup> More than 80% of Elsevier's revenues are derived from electronic products and services, much of which we invest into developing new digital capabilities and in collaborations throughout the sector.

<sup>3</sup> <http://www.elsevier.com/online-tools/research-intelligence>.

<sup>4</sup> <http://www.elsevier.com/online-tools/research-intelligence/products-and-services/pure>.

<sup>5</sup> <http://www.elsevier.com/online-tools/research-intelligence/resource-library/resources/international-comparative-performance-of-the-uk-research-base-2013>.

<sup>6</sup> [www.snowballmetrics.com](http://www.snowballmetrics.com).

<sup>7</sup> <http://documents.worldbank.org/curated/en/2014/03/19260159/state-stem-research-sub-saharan-africa>.

<sup>8</sup> <http://www.uis.unesco.org/Library/Documents/higher-education-asia-graduate-university-research-2014-en.pdf>.

<sup>9</sup> <http://www.oecd.org/sti/researchers-on-the-move-the-impact-of-brain-circulation.pdf>.

<sup>10</sup> <http://www.elsevier.com/online-tools/research-intelligence/resource-library/resources/international-comparative-performance-of-the-welsh-research-base-2013>.

<sup>11</sup> <http://www.elsevier.com/online-tools/research-intelligence/resource-library/resources/comparative-benchmarking-of-european-and-us-research-collaboration-and-researcher-mobility>.

3. **Universities:** EuroStemCell and Kyoto University's Institute for Integrated Cell-Material Sciences (iCeMS)<sup>12</sup>; Universities of Kyoto, Tokyo, Fudan, Manchester, and Michigan; Northwestern University; Johns Hopkins University; La Sapienza.
4. **Businesses:** Samsung; Siemens; Astra Zeneca; Novo Nordisk; Sanofi; Forum 48<sup>13</sup>; Inno360; MedImmune; Michigan Corporate Relations Experts portal network.
5. **Researchers:** researchers from around 3,000 institutions worldwide access the portfolio to demonstrate their performance to funding bodies, and to promotion and tenure panels.

We would like to participate in a workshop or event to discuss the value of using metrics in research assessment and showcasing of performance, and offer our expertise to HEFCE to contribute to determining the optimal approach to meet their objectives.

## B. For what purposes are research metrics used?

A **metric is a numerical measurement** that provides quantitative information about, in this document, research performance. Metrics are used by those involved in research as one of the inputs in making decisions. Research is growing<sup>14</sup> and becoming more complex<sup>15</sup>, and **metrics provide digestible insights from the massive amounts of data that these activities generate**. Metrics generation can be automated, and when used at scale can help to flag change early, such as the emergence of new disciplines.

In general, there are four types of situations in which research metrics are used:

1. **Evaluating research performance** is typically carried out by those in a position of authority relative to the entity(ies) being evaluated. This activity results in a funding body or university administration allocating funds, a dean or department head deciding which researcher to recruit or award tenure to, or a PhD student selecting a professor that they consider offers them the greatest chance of gaining a strong doctorate, for example.
2. **Showcasing performance** is generally conducted by those who are competing for limited resources. For instance, researchers look for ways to demonstrate that they should receive funding in preference to others, a research group leader may showcase their past performance to help to secure additional time on the Large Hadron Collider, or university administration may highlight areas of outstanding performance to help to attract overseas students to their institution.
3. **Scenario modelling of potential outcomes**, sometimes referred to as "fantasy football". Examples are a research group leader wondering how recruiting Dr A rather than Dr B would impact their team's performance, a dean considering the financial implications of combining linguistics and language, or university administration considering how to restructure the physics and chemistry schools.
4. **Ranking from high to low**. A parent supporting their child in university applications may want to know which university is the best, and a researcher might ask where they can find the leading research group in the area of photonics, for example.

---

<sup>12</sup> <http://www.elsevier.com/online-tools/research-intelligence/resource-library/resources/stem-cell-research-trends-and-perspectives-on-the-evolving-international-landscape>.

<sup>13</sup> <http://www.forum48.org/chicago-program-and-presentations/>; see presentation of Dr Holly Falk-Krzesinski.

<sup>14</sup> The number of researchers globally increases by about 1% per year from the current estimate of 7 million.

From International Comparative Performance of the UK Research Base – 2013, page 22, <http://www.elsevier.com/online-tools/research-intelligence/resource-library/resources/international-comparative-performance-of-the-uk-research-base-2013>.

<sup>15</sup> Researchers increasingly collaborate outside their universities and countries, resulting in an ever more complex network of research activity; trends visible in SciVal, <http://www.elsevier.com/online-tools/research-intelligence/products-and-services/scival>.


A metric in isolation is not especially useful in providing insights; one cannot evaluate whether an Awards Volume<sup>16</sup> of £3m represents an activity to which more funds should be allocated, or whether showcasing an Intellectual Property Volume<sup>17</sup> of 41 patents would be particularly attractive to overseas research students. For this reason, **those using metrics typically view them in the context of other metrics, by means of benchmarking and / or ranking.** Their distinguishing characteristics are summarised in [Table 1](#).

[Table 1](#): the characteristics of benchmarking and ranking

Characteristic	Benchmarking	Ranking
Currency of insights	Real time	Snapshot, usually annual
Selection of metrics used	Flexible selection and calculation options <sup>18</sup> per question	Fixed
Manner of use of metrics	Individual, separate metrics	Aggregated metrics that often have weighted contributions to the rank
Intended use	Private within a personal network	Public
Nature of comparison	Relative performance	Absolute performance
Situations	Evaluating Showcasing Scenario modelling	Ranking

**Benchmarking** involves the selection of metrics that are suitable to help to answer a particular question, and different metrics will likely be selected for different questions. The results are for the use of that individual, and perhaps by their network, and can be generated from current data whenever the questions arise.

**Ranking** is conducted by bodies who each determine a fixed method that is applied to a frozen snapshot of data. Metrics and sometimes opinions are usually aggregated into a single number that represents performance across multiple activities, and may have weighted contributions to the final aggregated rank: rankings are sometimes criticised for not having transparent methodologies.

<sup>16</sup>  Awards Volume: a Snowball Metric measuring volume of awards granted and available to be spent.

<sup>17</sup>  Intellectual Property Volume: a Snowball Metric measuring volume of patents and licenses.

<sup>18</sup> Examples of calculation options are limiting the analysis to awards from the Medical Research Council, or focusing only on original research articles or conference proceedings but not on reviews and editorials.

## C. Guiding principles for use of metrics in research assessment

Elsevier has extensive interactions with policy-makers, funding bodies, universities, businesses and researchers through our global networks. We have used this input to derive and test the following Principles that embody our best practise in the use of metrics in research performance assessment.

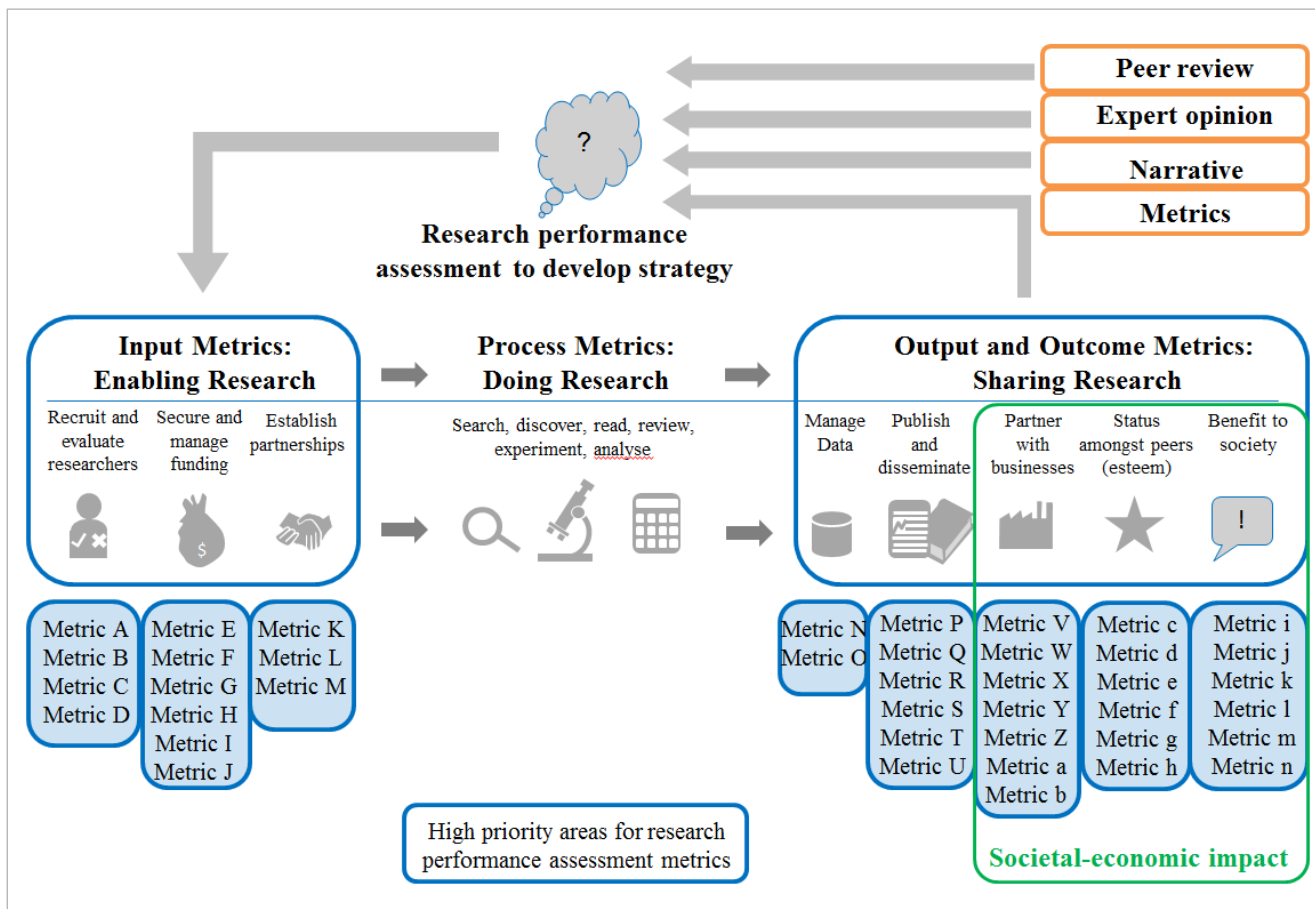
### 1. Principle 1: a set of multiple metrics distributed across the entire research workflow is needed.

- a. Metrics are used by those involved in research as one of their inputs in making decisions. These **decisions relate to many diverse questions, and encompass a wide variety of activities**, such as securing funding, collaboration, and sharing the results of research for others to build on by publishing articles, and by depositing the research data underpinning those articles.
- b. Even **within one area of activity, a question may be addressed by a variety of perspectives**. The extent of engagement of academia with business, for instance, has different meanings in different situations and to different users: metrics on the amount of cross-sector collaboration, consultancy, patents, number of FTEs employed by spin-outs, and so on, are needed.
- c. It is important to be able to quantify the value that research offers to those who consume it but who are not likely to build on it by applying for awards, or by citing it when they publish, such as the general public and clinicians. These contributions are generally referred to as “**societal and economic impact**” and comprise partnerships with businesses, status amongst one’s research peers (“esteem”), and benefit to society beyond the world of research.
- d. Elsevier, working in close partnership with the sector, has made **significant progress in implementing the vision of multiple metrics being available across the entire research workflow**<sup>19</sup> ([Figure 1](#)). We divide these activities into: “inputs” or the resources with which research is conducted; “processes” or the act of doing research; and “outputs and outcomes” which are the deliverables of the research.

---

<sup>19</sup> This is also documented in the literature. See, for example, H.F. Moed and G. Halevi, The Multidimensional Assessment of Scholarly Research Impact, accepted for publication by the *Journal of the Association for Information Science and Technology*, and available at <http://arxiv.org/abs/1406.5520>.

Figure 1: vision of a set of metrics across the research workflow to support optimal decision-making



2. Principle 2: metrics must be available to be selected for all relevant peers.

- a. **Benchmarking and ranking require peers** against which to judge performance as above or below average: knowing that a researcher reviewed 4 grants is meaningless without context.
- b. In answer to any one question, such as “In which areas am I excellent?” **different users will almost certainly select different peers against which to benchmark.** A university may well have a standard set of peers whose performance they monitor, but these will likely differ from the peers monitored by molecular biologists, and again by economists.
- c. The same users may also wish to **select different sets of peers when asking different questions**; economists might choose a distinct set of comparators to help decide to whom to award tenure, to those they would select to help to identify their global areas of excellence.

3. Principle 3: the generation and use of metrics should be automated and scalable.

- a. **Research performance is dynamic and constantly changing.** What was true 5 years ago may not remain so today.
- b. Also, **possibilities of looking at performance evolve** with the expansion of data sources and metrics available, necessitating the periodic revalidation of the evaluation methodology.
- c. These considerations, together with Principles 1 and 2 above, necessitate that **metric calculation can be automated so as to be carried out as needed on the most up-to-date version of the dataset, and is scalable** to provide context for any and all peers of interest.



- d. This is most likely achieved by the research community who define the metrics (Principle 9, below) **partnering with commercial providers of research information**, as has been the case in the generation of the free-of-charge, institution-endorsed, and technically feasible Snowball Metrics<sup>20</sup>.

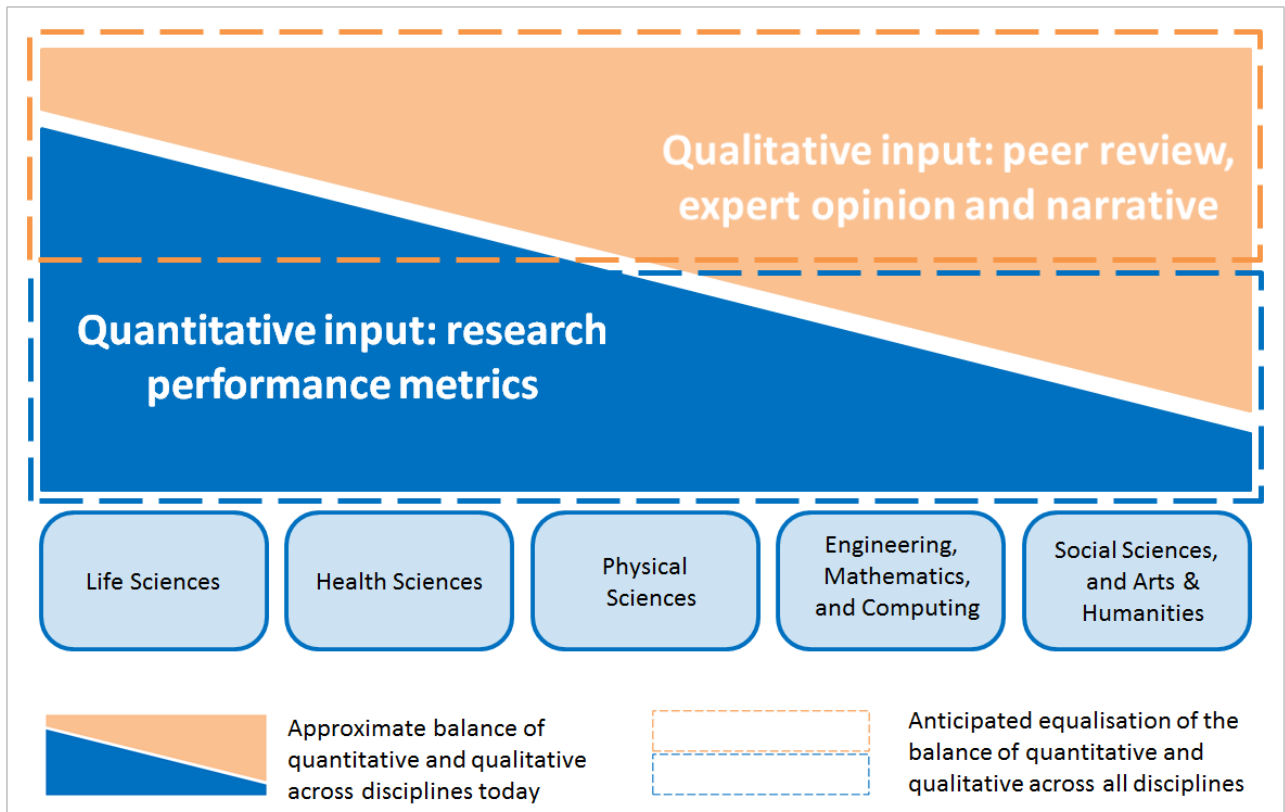
**4. Principle 4: quantitative information provided by metrics must be complemented by qualitative evidence to ensure the most complete and accurate input to answer a question.**

- a. Types of qualitative information are:
  - i. **Peer review**, which is the formal evaluation of a researcher's inputs, outputs or outcomes by those with competence in a discipline similar to that of the researcher. It generally results in a written appraisal, of which at least two are typically secured per input, output or outcome.
  - ii. **Expert opinion**, which is the more informal consultation of discipline experts, and/or those who know a researcher well. It generally results in a discussion.
  - iii. **Narrative**, which is a descriptive "story". Benefit to society, often referred to as "impact", may depend on so many discrete outputs, produced over such a long period and from so many researchers, institutions and countries that it is very difficult or even impossible to be certain that this can always be detected from the data. These situations are best served by "case studies".
- b. **Qualitative information would explain why, for instance, a metric is revealing a gap in a researcher's output.** Human input is needed to know whether the gap is due to having a child, family illness, a career break, moving to a new laboratory, or working with an industrial partner on confidential research that cannot be published, for example.
- c. The balance between the inputs of these approaches to an assessment differs depending on the particular question, and on the disciplines involved in the assessment. Currently, for instance, qualitative inputs tend to be more heavily relied on in social sciences and arts & humanities than they are in life sciences, although there are indications that the balance between disciplines is becoming increasingly similar (Figure 2). Regardless of the extent of this balance, **the critical point is that both quantitative and qualitative inputs should be used as inputs into making decisions, and ignoring any of these four approaches will lead to important information being missed.**

---

<sup>20</sup> Snowball Metrics recipes are available at [www.snowballmetrics.com/metrics](http://www.snowballmetrics.com/metrics).

Figure 2: model of the current and future balance of qualitative and quantitative inputs to research performance assessment amongst disciplines




5. **Principle 5: the combination of multiple metrics gives the most reliable quantitative input.**

- a. **Every metric has weaknesses, and there is no single perfect metric to give the complete quantitative picture in answer to any question.** Just as peer review generally relies on appraisals from at least two experts in the field, so does optimal quantitative information require input from at least two metrics, and preferably more, with metrics based on distinct data sources giving the richest picture.
- b. **Access to the broad set of metrics illustrated in Figure 1 is therefore critical;** as a figurative example, input into the societal-economic benefit of research would be stronger using both metrics V and W, than either alone, and strongest when using metrics V, W, d, f, k and m.

6. **Principle 6: disciplinary and other characteristics that affect metrics, but that do not indicate different levels of performance, must be taken into account.**

- a. **“Non-performance variables”, such as discipline, size and age may affect metrics, and may need to be taken into account** when comparing entities that comprise a mixture of these variables, such as a selection of institutions, or early career with established researchers.
  - i. **Discipline:** researchers in life sciences tend to publish more often, work with more co-authors, and refer to more previous work, than those in mathematics, for example, and will tend to receive higher values of metrics like Citations per Output<sup>21</sup> and Awards

<sup>21</sup>  Citations per Output: a Snowball Metric measuring average citations received per item of scholarly output.




Volume per FTE<sup>22</sup>. This does not necessarily indicate better performance, but merely reflects disciplinary behaviour. However, if these behavioural characteristics are not taken into account, they can skew apparent performance.

- ii. **Size:** larger entities will tend to have higher metrics when size is a factor, such as with total Citation Count<sup>23</sup>, or Market Share<sup>24</sup>.
  - iii. **Age:** older outputs have had more time to attract attention, and to receive higher counts of interest. In addition, activities such as citations and online activity take longer to accumulate in social sciences than in health sciences, for example.
- b. **Non-normalised metrics are very useful in some situations, and must be retained within the set available.** They tend to be more straightforward and transparent than normalised metrics, lending themselves more easily to validation. This makes them equally powerful tools for both evaluating and showcasing performance.
- c. Nevertheless, **the set of metrics available (Figure 1) must offer the option to normalise.** This may be addressed by one or both of:
- i. **The metrics themselves.** For example, Field-Weighted Citation Impact<sup>25</sup>, and Source-Normalised Impact per Paper (SNIP)<sup>26</sup> are normalised for both discipline and age. Number of Patents Granted per FTE<sup>27</sup> is normalised for size.
  - ii. **Functionality built into the tools that display the metrics.** For example, a filter that limits an entity to a particular discipline can enable the use of non-normalised metrics to compare entities.


## 7. Principle 7: metrics should be carefully selected to ensure that they are appropriate to the question being asked.

- a. **A research group leader considering which newly graduated PhD student to recruit to their team** would do best to select metrics that do not rely on high numbers of citations, since little time for citations to accumulate will have passed in the student's career; *h*-index<sup>28</sup>, for instance, would be a poor choice to look at the performance of a researcher so early in their career. The research group leader would gain better insights from International Collaboration<sup>29</sup> or the altmetric Scholarly Activity<sup>30</sup>, which give useful information as soon as an output is published.


<sup>22</sup>  Awards Volume per FTE: a Snowball Metric measuring the average volume of awards granted and available to be spent per FTE.


<sup>23</sup>  Citation Count: a Snowball Metric measuring total citations received by scholarly output.


<sup>24</sup>  Market Share: a Snowball Metric measuring percentage of sector total research income per institution.


<sup>25</sup>  Field-Weighted Citation Impact: a Snowball Metric measuring actual citation count relative to the expected world citation count.

<sup>26</sup> Source-Normalised Impact per Paper (SNIP): a free journal metric that can be considered a field-normalised version of the Impact Factor, [www.journalmetrics.com](http://www.journalmetrics.com).

<sup>27</sup>  Number of Patents Granted per FTE: an option within the Snowball Metric "Intellectual Property Volume" that measures the average volume of patents and licenses per FTE.

<sup>28</sup>  *h*-index: a Snowball Metric measuring the balance between productivity and citation impact.

<sup>29</sup>  International Collaboration: a Snowball Metric measuring proportion of internationally co-authored scholarly outputs.

<sup>30</sup>  Scholarly Activity: a Snowball Metric measuring the number of times outputs have been shared in online tools that are typically used by academic scholars, such as ResearchGate and Mendeley.

- b. As highlighted in the San Francisco Declaration on Research Assessment<sup>31</sup>, a researcher's performance should not be assumed to be equal to the Impact Factor, or an equivalent, of the serials they publish in. The Impact Factor, or equivalent, represents the average performance of a serial, and any output is more likely to perform above or below average, than at average. **Metrics that measure the actual performance of the researcher's output are preferred.**


**8. Principle 8: we cannot prevent the inappropriate or irresponsible use of metrics, but we can encourage responsible use by being transparent, and intolerant of "gaming".**

- a. **Metrics will always be used in a way that some consider inappropriate or irresponsible.** It is not advisable to try to combat this within the tools used to generate and visualise metrics, because the diversity of questions typically means that trying to address one such instance will create a limitation in the ability to answer other questions. The preferred approach is to provide guidance as to best practise, such as that in the SciVal Metrics Guidebook<sup>32</sup>, and transparency on the data underlying the metrics.
- b. **The extreme version of inappropriate behaviour is the "gaming" of metrics to improve apparent performance.** Gaming, such as abuse of the normal practise of self-citation to increase apparent citation impact, is rare, and can remain so by:
- i. **Defining metrics to structurally limit the possible effects of gaming;** for example, it is difficult to fabricate overseas co-authors to artificially boost International Collaboration<sup>33</sup>, and the SciMago Journal Rank (SJR)<sup>34</sup> discounts journal self-citations that account for more than one-third of the total.
  - ii. **Providing functionality in the tools that visualise metrics** so that users can judge whether gaming is taking place, such as the option to discount self-citations.
  - iii. **Adopting best practice** in the use of metrics in research performance assessment:
    - a. Always use at least two metrics (Principle 5, above), since every metric has weaknesses that can be complemented by the strengths of others. Addressing any given question by using **multiple metrics is also likely to reduce gaming**, since it is less obvious which behaviour should be focused on to "win".
    - b. **Always use metrics alongside the qualitative inputs** of peer review, expert opinion and narrative. If one input gives a very different message from the others, then this is a red flag that warrants further investigation.
  - iv. Continuing to **view "gaming" with disdain and as a negative, harmful activity** that will not be tolerated.

---

<sup>31</sup> <http://am.ascb.org/dora/files/SFDeclarationFINAL.pdf>.

<sup>32</sup> <http://www.elsevier.com/online-tools/research-intelligence/resource-library/resources/scival-metrics-guidebook>.

<sup>33</sup>  International Collaboration: a Snowball Metric measuring proportion of internationally co-authored scholarly outputs.

<sup>34</sup> SciMago Journal Rank (SJR): a free journal metric that can be considered a field-normalised version of the Impact Factor, [www.journalmetrics.com](http://www.journalmetrics.com).

9. **Principle 9: those in the research community who apply metrics in their day-to-day work, and who are themselves evaluated through their use, should ideally define the set of metrics to be used. It is highly desirable that this same community, or those empowered by the community on their behalf, maintains the metric definitions.**

- a. This ensures that the **research community articulates its needs and endorses the solution, since the solution has been developed by itself rather than imposed**. It seems logical that this approach leads to metrics that are available and accessible to everyone with an interest in research performance, and that they are as simple and easy to implement as possible while still providing valuable information; the simplicity must sometimes be forfeited to address another strong need, as in the case of some normalised metrics (Principle 6, above).
- b. This principle also promotes a **pragmatic approach that uses the available metrics to get as close as possible to the objectives being addressed**, while ensuring that perfection does not become the enemy of good, and cause inertia.

10. **Principle 10: there should be no methodological black boxes.**

- a. Metrics are used as an input into resource allocation and recruitment. This has an effect on the direction of research, and on careers. It is therefore **essential that the underlying methods are open to build trust, and to support debate and improvement** when needed.
- b. **All calculation methods must be supplier-agnostic: freely available, transparent, and not proprietary**. We believe that discussions of “Elsevier metrics”, “Thomson metrics” and “Academic Analytics metrics”, for example, are not especially useful because the community does not benefit from competition at the level of metrics methodologies. Competition should instead occur at the levels of the data and infrastructure underpinning the metrics, and of the tools and services into which they are built; consequently, we at Elsevier openly share the methods we use to generate metrics, for example in the SciVal Metrics Guidebook<sup>35</sup>.

11. **Principle 11: metric methodologies should be independent of the data sources and tools needed to generate them, and also independent of the business and access models through which the underlying data are made available.**

- a. Metrics give the most **useful input into decision making when they are based on the most complete sets of data available**, so that they give the most complete quantitative picture possible. A policy-maker, funding body, university, business or researcher should be able to draw on all data sources available to them to generate metrics.

12. **Principle 12: aggregated or composite metrics should be avoided.**

- a. Aggregated metrics, of the type that tend to be used to generate global rankings, **reduce the transparency of the outcome** (Principle 10, above).
- b. Furthermore, the individual metrics generally tend to have a weighted contribution to the aggregate that is not always transparent, and these **weightings exaggerate the effect of one or a few metrics**, that others consider to be the most important, on the outcome.

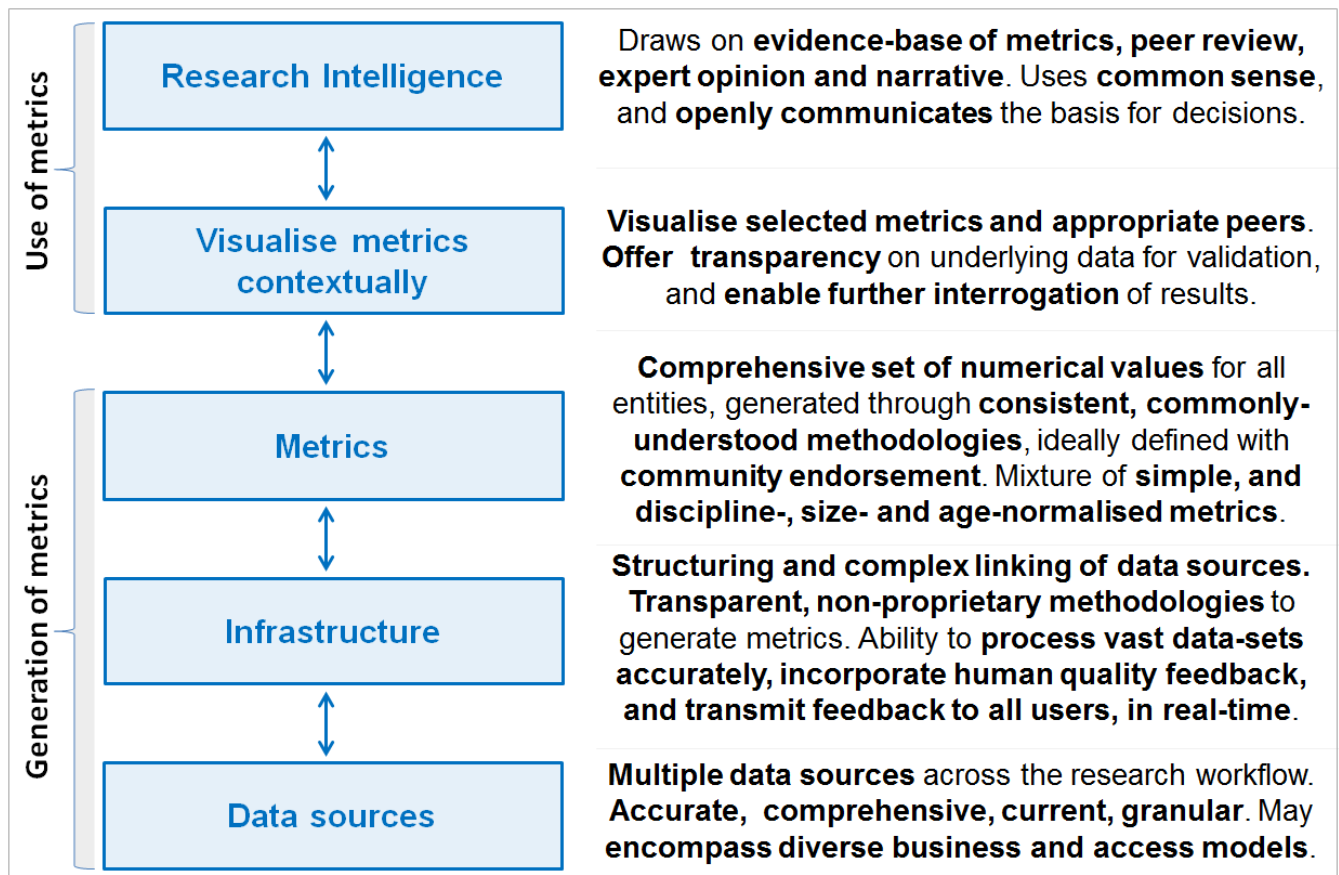
---

<sup>35</sup> <http://www.elsevier.com/online-tools/research-intelligence/resource-library/resources/scival-metrics-guidebook>.

## D. A model for generating and using metrics in research assessment

The use of metrics in research performance assessment clearly requires a multi-faceted approach. In section B, we have discussed the purposes for which research metrics are used, and in section C we have elucidated 12 principles that guide Elsevier's use of metrics in research performance tools. We have drawn these guidelines and observations together in the model shown in [Figure 3](#).

[Figure 3](#): a model of the elements needed to generate and use metrics in performance assessment



Metrics can only be generated from **data sources combined within an infrastructure that**:

1. **Organises, and links these data sources.** The infrastructure must link elements within a single data source – for example, publications in Elsevier's Scopus<sup>36</sup>, the largest abstract and citation database of peer-reviewed literature, are linked through multiple tags, such as author, affiliation, affiliation type, citing publications, publication type, journals, and discipline – as well as linking equivalent elements between different data sources – for example, the subset of a researcher's outputs that have been funded by a particular grant.
2. **Applies the metrics methodologies to the entities for which the user wishes to view them.** Any given metric can usually be calculated for multiple entities - individual researchers, groups of researchers, institutions, groups of institutions, disciplines, journals, countries, regions, and so on. The infrastructure must be able to “slice and dice” data sources flexibly, and, having extracted the appropriate data elements, process them in real time.

<sup>36</sup> <http://www.elsevier.com/online-tools/scopus>.

3. **Can incorporate feedback from manual quality checks into the data links and tags, and transmit these improvements to all users in real time.** The data elements being organised are produced by humans, and unavoidably contain incomplete and / or incorrect information, so that although automation can approach 100% accuracy, it can never be achieved. Underlying data quality is a critical component of using metrics, and can only be reached by acting on human inputs on the automated structure.

## E. Implications for using metrics in research assessment

Elsevier believes that the points made above can best be actualised by:

1. **The endorsement of a single set of metric methodologies as a standard that can be applied to various data sources, tools, and business and access models. We endorse Snowball Metrics<sup>37</sup>** which have informed our guiding principles, and we encourage HEFCE to consider taking this step.
2. **Working with commercial suppliers of research information to ensure the scalability and reproducibility of metrics generated as input into to research performance evaluation.** Suppliers are best positioned to provide the ecosystem – data, infrastructure, and tools to select metrics and peers - within which metrics can be most effectively generated and utilised.

## F. Response to the specific questions posed by HEFCE

### Identifying useful metrics for research assessment:

#### **What empirical evidence (qualitative or quantitative) is needed for the evaluation of research, research outputs and career decisions?**

Robust evaluation relies on a combination of qualitative information – peer review, expert opinion and narrative – with the quantitative – metrics. See Principle 4, section C, for a more complete response.

#### **What metric indicators are currently useful for the assessment of research outputs, research impacts and research environments?**

Elsevier currently uses a range of metrics distributed across the research workflow ([Figure 1](#)). Snowball Metrics<sup>38</sup>, applied to the appropriate entities, are also a good representation of the scope of metrics needed.

#### **What new metrics, not readily available currently, might be useful in the future?**

The most noticeable areas of research workflow ([Figure 1](#)) where additional metrics are currently needed are:

1. Managing data, so as to share it in a discoverable and usable way for others. This area may cover the deposit of data, quantity of downloads indicating data reuse, and frequency of “citations” back to the data in new research publications, for example.
2. Esteem, or credit received from peers. This area may cover editorial and review activities for funding bodies and publications, conference attendance and organisation, and membership of learned societies, for instance.
3. Benefit to society (“impact”). This area may include media mentions, contributions to public policy, and contributions to best practice guidelines, for example.
4. Metrics in the areas of post-graduate education and / or teaching are also lacking.

---

<sup>37</sup> [www.snowballmetrics.com](http://www.snowballmetrics.com).

<sup>38</sup> [www.snowballmetrics.com/metrics](http://www.snowballmetrics.com/metrics).



### Are there aspects of metrics that could be applied to research from different disciplines?

Metrics can be calculated for entities in all disciplines: Awards Volume<sup>39</sup>, and Public Engagement<sup>40</sup>, can be calculated for arts & humanities as well as for chemistry. Whether metrics are used in particular disciplines, as well as in answering particular questions, is somewhat a matter of judgment:

1. Awards Volume may be of little interest in Arts & Humanities given the relatively small size of the funding typically being considered when compared to Life Sciences.
2. Data source coverage may also affect the decision to use a particular metric, even if the metric is in principle of interest; Scholarly Output<sup>41</sup> may not always be a useful metric in Arts & Humanities given the relatively poor coverage of the commercial databases in this discipline. This coverage is, however, improving all the time; around 5,000 books are added to Scopus every month, for example.
3. An appropriate choice of peers against which to benchmark may enable use of a metric; for instance, Scholarly Output may be a useful way to compare a selection of institutions with a similar degree of focus on Arts & Humanities, since the database coverage will affect all institutions equally.

### What are the implications of the disciplinary differences in practices and norms of research culture for the use of metrics?

Characteristics of researcher behaviour differ between disciplines, and can affect the metrics and skew apparent performance. The set of metrics must take this into account, as well as metrics that normalise for size and age. However, non-normalised metrics are also valuable components of the set. See Principle 6 in section C for a more complete response.


### What are the best sources for bibliometric data? What evidence supports the reliability of these sources?

Commercial databases with stable content, rigorous peer reviewed inclusions, and transparent coverage policies are the best sources; in practice, this means Elsevier's Scopus or Thomson Reuter's Web of Science, and not Google Scholar which has fluid and unreviewed content.

Scopus<sup>42</sup> is the largest abstract and citation database of peer-reviewed literature, with 53 million records from 5,000 publishers; it also searches across over 25 million patents from five patent offices. The titles indexed are selected by the independent Content Selection and Advisory Board<sup>43</sup>, working with three local expert boards<sup>44</sup> in China, Korea and Thailand, and comprise almost 21,000 peer-reviewed journals including 2,800 open access journals, over 400 book series, over 30,000 stand-alone books, 5.5 million conference papers, and over 350 trade publications. As such, it has the broadest disciplinary and geographical coverage in the world<sup>45</sup>.

Scopus has been independently validated as the preferred source of bibliometric data in national research exercises in the United Kingdom, Australia, Italy, and Portugal. **Miguel Seabra, the President of Fundação para a Ciência e a Tecnologia (FCT)**, which runs the Portuguese national evaluation exercise, and future President of Science Europe, commented:

*“The bibliometric analyses conducted by Elsevier are an important component of the review process, providing very useful insights for the international review panels. Their involvement will thus contribute to the overall aim of this assessment exercise, which is to reinforce the role*

<sup>39</sup>  Awards Volume: a Snowball Metric measuring volume of awards granted and available to be spent.

<sup>40</sup>  Public Engagement: a Snowball Metric measuring attendees at public events.

<sup>41</sup>  Scholarly Output: a Snowball Metric measuring productivity.

<sup>42</sup> <http://www.elsevier.com/online-tools/scopus>.

<sup>43</sup> <http://www.elsevier.com/online-tools/scopus/content-overview#scopus-content-selection-and-advisory-board>.

<sup>44</sup> <http://www.elsevier.com/online-tools/scopus/content-overview#local-content-boards>.

<sup>45</sup> [http://www.elsevier.com/\\_data/assets/pdf\\_file/0006/155427/Scopus-Content.pdf](http://www.elsevier.com/_data/assets/pdf_file/0006/155427/Scopus-Content.pdf).



*of research units as a fundamental pillar in building a modern and competitive scientific system.*<sup>46</sup>

### What evidence supports the use of metrics as good indicators of research quality?

Some metrics are clearly very good indicators of research quality. For example, the 2013 report on the International Comparative Performance of the UK Research Base, produced for the Department of Business, Innovation and Skills, reports that the UK produced 15.9% of the world's most highly-cited publications over the period 2008-2012, and that it is the most efficient of its comparator countries in that it produced 3.9 publications and 43.1 citations per unit of gross domestic expenditure on research and development (GERD) in 2012<sup>47</sup>.

Further evidence is the growing demand for metrics and tools, as evidenced by the steep growth in use of the Elsevier Research Intelligence portfolio which integrates metrics into all of its workflows; the number of UK customers of Pure<sup>48</sup>, our research information management system, has increased from zero in 2008, to more than 20 in 2013. Benchmarking tools such as SciVal, InCites and Academic Analytics are becoming more and more widespread. There is also increasing engagement of the research community in metrics- and standards-generating efforts, such as Snowball Metrics<sup>49</sup>, euroCRIS<sup>50</sup> and CASRAI<sup>51</sup>.

Metrics will never convey the complete picture of research quality, however, and should always be used as a complement to the qualitative approaches of peer review, expert opinion and narrative (Principle 4, section C), and not as a substitute for human judgment.

### Is there evidence for the move to more open access to the research literature to enable new metrics to be used or enhance the usefulness of existing metrics?

Yes; the drive towards open access will enable the creation of some new metrics, such as usage of institutional repositories. However, metrics should be independent of the business and/or access models under which the research literature is published so that they are based on the most complete sets of data available, and give the most complete quantitative picture possible (Principle 11, section C).

### How should metrics be used in research assessment?

#### What examples are there of the use of metrics in research assessment?

See section A for examples of Elsevier customers using metrics in research assessment.

Scopus has been independently validated as the preferred source of bibliometric data in national research exercises in the United Kingdom, Australia, Italy, and Portugal; please refer to the response in this section to the question, "What are the best sources for bibliometric data? What evidence supports the reliability of these sources?" for a quote from Miguel Seabra, the President of Fundação para a Ciência e a Tecnologia (FCT), which runs the Portuguese national evaluation exercise, and future President of Science Europe, on Elsevier's involvement in the review process.

---

<sup>46</sup> <http://www.elsevier.com/about/press-releases/science-and-technology/national-funding-agency-for-science-and-technology-in-portugal-selects-elseviers-analytical-services-to-support-evaluation-of-r-and-d-units>.

<sup>47</sup> <http://www.elsevier.com/online-tools/research-intelligence/resource-library/resources/international-comparative-performance-of-the-uk-research-base-2013>. The figure on highly-cited publications is reported in Figure 1.1A, and those on output and citations per unit GERD in Figure 6.1.

<sup>48</sup> <http://www.elsevier.com/online-tools/research-intelligence/products-and-services/pure>.

<sup>49</sup> Snowball Metrics aims to develop the international standard for metrics used to build and monitor institutional strategies; [www.snowballmetrics.com](http://www.snowballmetrics.com).

<sup>50</sup> euroCRIS is a non-profit organisation interested in research information systems and their inter-operability; [www.eurocris.org](http://www.eurocris.org).

<sup>51</sup> Consortia Advancing Standards in Research Administration Information (CASRAI) is a non-profit standards development organisation; [www.casrai.org](http://www.casrai.org).

### **To what extent is it possible to use metrics to capture the quality and significance of research?**

The degree to which metrics reflect research activities is growing, as the available data sources and metrics expand; for example, the recent surge of interest in Altmetrics<sup>52</sup> indicates that the sector is interested in exploring new data sources which add to the comprehensiveness of the picture that research metrics provide.

Metrics will however never reflect 100% of research activity, quality and significance, and should not be used as though they do. They give the most reliable insights into research performance when used together with, not instead of, the qualitative inputs of peer review, expert opinion and narrative (Principle 4, section C).

### **Are there disciplines in which metrics could usefully play a greater or lesser role? What evidence is there to support or refute this?**

See Principle 4 in section C for a more complete answer.

Metrics currently play a greater role in life sciences, than in arts & humanities, with the contribution of peer review, expert opinion and narrative being higher in the latter ([Figure 2](#)), as illustrated by the differing use of metrics by the REF panels. However, the balance is shifting to become more even across all disciplines due to improvements in data sources and metrics, and experience in using metrics in performance assessments.

### **How does the level at which metrics are calculated (nation, institution, research unit, journal, individual) impact on their usefulness and robustness?**

The majority of metrics can be calculated for multiple entities – nations and groups of nations such as continents and socio-political networks such as the rapidly growing BRIC<sup>53</sup> countries, institutions and groups of institutions such as the Russell Group, researchers, groups of researchers such as teams and international collaboration networks, granular sets of items such as the outputs funded by a particular grant, and disciplines.

Care should be used when using metrics to investigate the performance of small entities, like researchers, because missing elements in the data source may have a greater impact on apparent performance. Careful use of metrics in these situations remains a useful complement to peer review, expert opinion, and narrative, however.

### **‘Gaming’ and strategic use of metrics:**

#### **What evidence exists around the strategic behaviour of researchers, research managers and publishers responding to specific metrics?**

There are well-known examples, such as the strategic allocation by research managers of researchers to Units of Assessment; the publication by researchers of more papers, each containing a smaller amount of work, when productivity is measured (“salami slicing”); and the excessive citation by editors of content within their journal to boost Impact Factors. It should be noted that these practices are rare, but examples tend to gain notoriety which gives the impression that they are more widespread than they actually are.

The disdain with which these activities are viewed, however, tends to lead to self-correcting mechanisms which limit such behaviours such that they remain exceptions.

#### **Has strategic behaviour invalidated the use of metrics and/or led to unacceptable effects?**

Strategic behaviour has led some to lack trust in the use of metrics. However, it has also driven improvements: some metrics have evolved to become more resilient to strategic behaviour; tools offer options to check for unscrupulous actions, such as removing self-citations from a calculation; and the users of metrics have become more aware of and vigilant as regards unacceptable behaviour.

---

<sup>52</sup>  Altmetrics: Snowball Metrics measuring counts of online activity stimulated by Scholarly Output.

<sup>53</sup> BRIC countries: Brazil, Russia, India, and China.

**What are the risks that some groups within the academic community might be disproportionately disadvantaged by the use of metrics for research assessment and management?**

Those with relatively poor coverage in the commercial databases are currently disadvantaged, such as the social sciences and arts & humanities. However, trends are still apparent in the data for these fields, such that metrics remains a useful complement to the qualitative inputs of peer review, expert opinion and narrative in these areas (Principle 4, section C).

**What can be done to minimise ‘gaming’ and ensure the use of metrics is as objective and fit-for-purpose as possible?**

Several options are open to reduce “gaming” in the use of metrics in performance assessment. These are described in Principle 8 of section C.

Submitted on behalf of Elsevier by

**Dr Lisa Colledge**

**Head of Research Metrics**

[l.colledge@elsevier.com](mailto:l.colledge@elsevier.com)