

Global movement to reform researcher assessment gains traction

One aim is to recognize a wider range of research contributions.

A growing global movement toward holistic approaches to evaluating researchers and research aims to value a broader range of contributions than an institute's reputation and such metrics as numbers of publications in high-impact journals, citations, and grant monies. Contributions that go largely unrewarded include committee service, outreach to the public and to policymakers, social impact, and entrepreneurship.

An early push was the San Francisco Declaration on Research Assessment in 2013. DORA has grown into a worldwide initiative for which reducing the emphasis on journal impact factor has been a “hobbyhorse,” says program director Zen Faulkes. “But we are broadening our efforts in assessment reform.” As of September, more than 20 000 individuals and about 3000 organizations in 164 countries had signed DORA.

A related effort spearheaded by the European Commission, the European University Association, and Science Europe—an association of funding agencies that spends more than €22 billion (roughly \$24 billion) annually—is widely seen as having the most punch. In July 2022 they laid out guiding principles for reform, and in December 2022 they established the Coalition for Advancing Research Assessment (CoARA). More than 600 universities, funders, learned societies, and other organizations, overwhelmingly in Europe, had signed on as of late August. Signatories commit to examining their research assessment procedures within a year and to trying out and reporting on alternative approaches within five years.

CoARA is different from earlier assessment reform statements, says Sebastian Dahle, a physicist at the University of Ljubljana in Slovenia, a CoARA signatory. “In signing, organizations have to create an action plan. The agreement drives things forward. It's not legally binding, but it keeps people engaged.”



OLEKSANDR BEREZKO

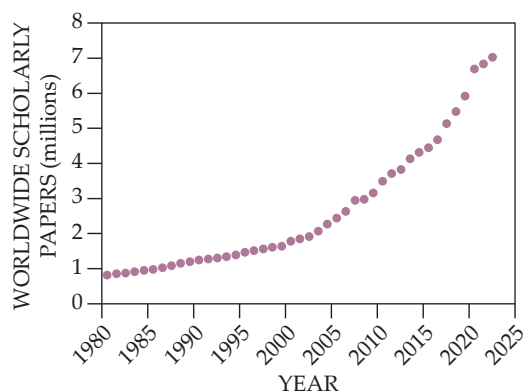
KAREN STROOBANTS (on screen), vice chair of the Coalition for Advancing Research Assessment steering board, at the 2023 annual conference of the European Council of Doctoral Candidates and Junior Researchers in Uppsala, Sweden. The council's president, Sebastian Dahle, is seated, third from left. The other panelists represent other organizations that also support reform of research assessment.

The motivation to reform research assessment stems largely from frustration with the publish-or-perish culture that has developed in recent decades, and the movement's aims represent a return to earlier norms. Assessing researchers “almost entirely” on the quantity and citations of their publications “creates poor incentives,” says Elizabeth Gadd, vice chair of CoARA's steering board. It leads to scholars “salami slicing” to pad their publication counts, selling authorship, and committing fraud or misconduct, she says, adding that it's “hugely

problematic” that publications are so central to evaluating researchers.

In Europe, says Toma Susi, who works on low-dimensional materials at the University of Vienna, “there is a widespread feeling that the academy has lost autonomy in how it evaluates researchers and institutes.” The ubiquitous impact factors are commercial and generated in “non-transparent” ways, he says.

Cassidy Sugimoto, chair of the school of public policy at Georgia Tech, studies scientometrics and inequalities in the scientific workforce. She says that the pres-



THE SOARING NUMBER of publications across all fields worldwide reflects the publish-or-perish culture prevalent in academia. It's among the drivers in the movement to reform research assessment. (Adapted from Digital Science, Dimensions software, available from <https://app.dimensions.ai/analytics/publication/overview/timeline>, accessed on 12 September 2023.)

sure to publish taxes the mental health of scholars. The entire research community is affected, but the burden tends to be higher on women and people of color, she says, in part because on average they are awarded less research funding and their papers are cited less often. If the current system of assessment leads to poorer mental health—and more attrition—she asks, “is it meeting the goals of scientists? Is it best for science?”

Another stress on mental health and a motivation for assessment reform is the frequency of evaluations. Between appraisals, tenure, promotions, prizes, and grant applications, “researchers are evaluated left, right, and center,” says Gadd, a specialist on research evaluation, scholarly communications, and research culture at Loughborough University in the UK. “It’s a personal driver for me to provide the best environment for researchers to do their research.”

Additionally, the open-science movement, which espouses making research—data, methods, software, and more—available to benefit the advancement of both science and society at large, is converging with efforts to reform research assessment. The emphasis on publishing in top journals leads to delays in sharing new discoveries, explains Johan Rooryck, executive director of the international open-access publishing initiative cOAlition S; the 28 funding agencies that belong to the coalition collectively invest about \$40 billion a year in research. Around

90% of submissions are rejected by top journals, in many cases without review. “It’s impossible that 90% of articles are bad science,” he says. Scholars submit their manuscripts repeatedly until they find a journal that accepts them. “It creates enormous waste.”

The extreme competition also threatens publishing’s peer-review process, says Rooryck, who as an editor-in-chief of the linguistics journal *Glossa* has witnessed firsthand the increasing difficulty of finding reviewers. As a contribution to their research communities, scholars have traditionally volunteered to peer-review papers. But people are practical, Rooryck says. “Without compensation or acknowledgment, why spend time writing reviews?” The system has to change, he says, otherwise it will “come to a screeching halt.”

“The way research is conducted has evolved a lot over the past two decades,” says Nicolas Walter of the European Science Foundation, which supports CoARA with infrastructure, staff, and financial management. He points to the sheer volume of data and to new ways of generating and sharing data. The outputs of research have also changed, he continues, “so the way we assess research has to evolve.” Or, as Susi, who participated in the drafting of the CoARA agreement, puts it, “the bottom line is that qualitative things require qualitative evaluation.”

Committing to reform

CoARA sets out four core commitments to guide the reform of research assessment:

- Recognize diversity in the contributions to, and careers in, research.
- Base research assessment primarily on qualitative evaluation, for which peer review is central, supported by responsible use of quantitative indicators.
- Abandon the inappropriate uses in research assessment of journal- and publication-based metrics, in particular the inappropriate uses of journal impact factor and *h*-index.
- Avoid the use of rankings of research organizations in research assessment.

Additional supporting commitments include agreeing to allocate resources, raise awareness, and share results from reform experiments.

CoARA is intended to provide guid-

ance, not prescribe actions. Signatories sign on to the principles, but have to find ways to apply them that work in their specific settings. Research cultures and needs vary by discipline, institution, and country.

This summer, CoARA launched 10 working groups to explore issues relevant to research assessment. Dahle, for example, who is president of the European Council of Doctoral Candidates and Junior Researchers, chairs a working group on early-career researchers. Rooryck chairs one that is looking at how to recognize and reward peer review. Another focuses on multilingualism and language biases in research assessment.

CERN was an early signer of CoARA. CERN’s practices and culture already align with CoARA principles, says Alexander Kohls, the lab’s group leader for scientific information services. “If you talk to a theorist at CERN, it doesn’t matter whether a paper appears in arXiv or a top journal; the content is valued, the venue less so,” he says. But, he adds, some CERN collaborators say things along the lines of, “I don’t want to make my research output open. I prefer to protect it for my own use.” Kohls says that the lab wants “to push forward research assessment reform” in order to nudge other institutions to “follow the spirit” of what CERN has been doing for a long time.

In Poland and other eastern European countries, funding and jobs were historically not based on merit but rather on connections and politics, says Emanuel Kulczycki, head of a research group on scholarly communication at Adam Mickiewicz University in Poznań and an adviser to Poland’s ministry of science. A legacy of communism, he says, is that universities remain under government control and the academic community is “eager to trust in metrics.” At the same time, he adds, considering social impacts of research is not new there.

In countries with research structures like Poland’s, says Kulczycki, reform has to get the nod from the government, “but the black box of evaluation should be designed by the academic community.” CoARA could be helpful, he adds, for mining ideas and steps for their implementation. He notes that acknowledging multilingualism is crucial in his country. “Using your own language plays an important role in popularizing science and in attracting students,” he says.

Independent of CoARA, in 2020

China instituted reforms along the same lines. Before that, the country had been known for rewarding scholars with cash bonuses for publishing in top international journals. The reforms include valuing a wider range of research outputs and relying on comprehensive peer-review evaluations, says Lin Zhang, a professor of information management at Wuhan University, editor-in-chief of the international journal *Scientometrics*, and an adviser to China's ministry of education on research assessment reform. The earlier incentives improved Chinese researchers' global visibility, she says, "but at a cost of research integrity for some researchers." With new research assessment guidelines, the hope is to focus more on "novelty, scientific value, research quality, research integrity, and societal needs," Zhang says. In reforming research assessment, "China shares the same motivation as the rest of the world."

The US is seen by some as lagging in the area of research assessment reform. That apparent lag can be attributed partly to the decentralized university system and multitude of funding sources, says Sugimoto. And whereas some countries explicitly require, say, a certain number of publications in top-tier journals for someone to get a promotion, in the US such requirements are not typically codified, she notes. "Many of our practices are implicit. That makes them harder to combat."

The European Science Foundation's Walter expects the movement will gain traction in the US and more broadly. CoARA is less than a year old, he notes. "We are now in a phase where we need to engage outside of Europe."

Catalysts, not panaceas

In recent years, a smattering of funding agencies, institutions, and countries have begun experimenting with assessment reforms. Some funding agencies are using lotteries to award grants. Some are putting caps on the number of grants that a given investigator can receive. The Luxembourg National Research Fund (FNR) is broadening the range of contributions for which it awards prizes. As examples, FNR program manager Sean Sapcaru points to a new award that recognizes outstanding mentorship and another that was changed from naming the best publication to rewarding an outstanding research achievement. Funding can be a "blunt tool to shape behavior," he says.

Across Europe, many institutes and funding agencies have begun introducing narrative CVs for job and grant applications; the Netherlands, Norway, and the UK are among the pioneers. For narrative CVs, scholars are asked to limit the number of publications they list to perhaps 5 or 10 and to discuss their relevance. Researchers are also invited to write about other germane contributions and to explain why they are a good candidate for the proposed project or job. "The idea is to provide room for candidates to point out their contributions that may not fit into a traditional CV," says Robbert Hoogstraat, project leader for the Dutch Research Council's Recognition and Rewards program. "Maybe they participated in an open-science activity or wrote an opinion article for a newspaper."

Luxembourg's FNR is both using narrative CVs and studying their efficacy and reception. The FNR asks scientists to produce a two-page CV with three sections: a personal statement related to the research for which they are requesting funding, a description of their professional path, and a discussion of relevant achievements and outputs. So far, says Sapcaru, survey results show that 70% of reviewers and nearly as many researchers view narrative CVs positively. Starting next year, the European Research Council will let applicants add narrative descriptions to their CVs and will give more weight to project proposals than to past achievements.

Capping the number of papers listed in a CV helps level the competition in terms of career stage, gender, and geography, say proponents. Narrative CVs could also make it easier to get funded in interdisciplinary research areas and to switch fields. "The narrative CV is not a panacea," says Frédérique Bordignon, a researcher at the École des Ponts Paris Tech who studies bibliometrics and research integrity. "But it can be a catalyst to find better ways to assess researchers."

Despite funders being uniquely positioned to leverage change, says Angela Bednarek of the Pew Charitable Trusts, they can hit walls. She leads the Transforming Evidence Funders Network, a group of 70 private and public funders that aim to increase the societal impact of their research investments. In response to calls for projects, Bednarek notes, some early-career researchers say, "You are asking me to invest time for something that doesn't get me

the publications I need to get tenure."

For example, Bednarek says, a project "might synthesize existing data for use by decision makers," rather than involve pathbreaking research. She points to using data about the physical conditions of the ocean to set fishing limits in response to climate change. "Funders need to think about how research is rewarded and incentivized so they can support relevant and timely research," she says. "We don't call it assessment reform, but it's the same thing."

Metric challenges

A common misinterpretation of CoARA is that proponents aim to do away with metrics. Responsible use of metrics, Bordignon says, should provide context. A statement such as "I have published five articles in the last 10 years and have supervised 10 doctoral students in that period" mitigates the impact of having a single number describe one's work, she says. And since *h*-index grows with the number of publications, indicating academic age would explain the disparity of the *h*-indices between early-stage and seasoned researchers. In addition, metrics used in aggregate can be helpful in comparing the outputs of large institutions or countries.

Among the concerns with reforming assessment are that qualitative assessment may be more time consuming and may be more subjective than relying on metrics. It is harder, proponents admit, especially during a transitional period as applicants and assessors become familiar with new procedures. But, they say, the increased time required for qualitative assessments could be offset by reducing the total number of evaluations conducted.

DORA's Faulkes says he understands people's concern that "if you take away numbers, it's backroom deals and patronage." But, he says, "We are not saying to abandon metrics entirely." And, Faulkes adds, DORA is joining the "research on research" community to assist in understanding the use of narrative CVs and other qualitative peer-review practices.

"Even coldhearted metrics are not free of biases," says Lynn Kamerlin, a chemistry professor who was involved in science policy in Europe before moving to Georgia Tech last year. Limited jobs and funding are a "zero-sum" game, she says. "Unless the underlying issue of hypercompetition is solved, everything else is a Band-Aid."

Toni Feder 