

AGREEMENT ON REFORMING RESEARCH ASSESSMENT

20 July 2022

"'Publish or perish' and metrics have led us into a blind alley. Let's start recognising the full breadth of value created by researchers."

MARC SCHILTZ,
President of Science Europe

An assessment of the new *Agreement on Reforming Research Assessment*

R-Quest seminar 24th november 2022

Gunnar Sivertsen

NIFU

The Agreement

- Developed in collaboration between Science Europe and the European University Association
- Facilitated and supported by the European Commission
- Published on 20 July 2022, officially opened for signature on 28 September
- Implementation will be promoted by the Coalition for Advancing Research Assessment (CoARA) – to be constituted on 1st December



Outline

1. Three contexts for research assessment
2. The purposes of the assessment
3. What should be assessed?
4. The four core commitments
5. What is said about "peer review", "indicators", and "metrics"?
6. The relevance of the Agreement in each context
7. The knowledge base
8. My assessment of the Agreement

Outline

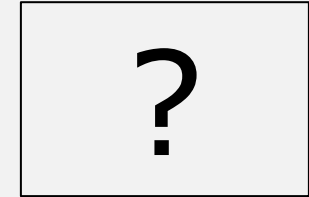
1. Three contexts for research assessment
2. The purposes of the assessment
3. What should be assessed?
4. The four core commitments
5. What is said about "peer review", "indicators", and "metrics"?
6. The relevance of the Agreement in each context
7. The knowledge base
8. My assessment of the Agreement

Three contexts for research assessment

1. “The assessment of **research performing organisations and research units**”
2. “The assessment of **research projects**”
3. “The assessment of **individual researchers and research teams**”

Direct responsibilities for the three contexts

1. Organizational evaluation



2. Project funding



3. Recruitment and promotion



The purposes of the assessment: Mainly allocation of limited resources in competitive contexts

1. Organizational evaluation

- “Allocating funding; public investment accountability; informing decisions on research priorities and improving the definition and implementation of research strategies”

2. Project funding

- “Allocating funding, informing project management and future research funding decisions, and making prize and award decisions”

3. Recruitment and promotion

- “Allocating funding, recruitment and hiring promotion, professional development review, and prize and award decisions”

What should be assessed?

- Changes in assessment practices should enable recognition of the *broad diversity* of:
 - valuable contributions that researchers make to science and for the benefit of society, including diverse outputs **beyond journal publications** and irrespective of the language in which they are communicated;
 - **practices** that contribute to **robustness, openness, transparency, and the inclusiveness** of research and the research process including: peer review, teamwork and collaboration;
 - **activities including** teaching, leadership, supervision, training and mentoring.

Outline

1. Three contexts for research assessment
2. The purposes of the assessment
3. What should be assessed?
4. The four core commitments
5. What is said about "peer review", "indicators", and "metrics"?
6. The relevance of the Agreement in each context
7. The knowledge base
8. My assessment of the Agreement

The four core commitments

1. Recognise the diversity of contributions to, and careers in, research in accordance with the needs and nature of the research
2. Base research assessment primarily on qualitative evaluation for which peer review is central, supported by responsible use of quantitative indicators
3. Abandon inappropriate uses in research assessment of journal- and publication-based metrics, in particular inappropriate uses of Journal Impact Factor (JIF) and h-index
4. Avoid the use of rankings of research organisations in research assessment

The four core commitments

1. Recognise the diversity of contributions to, and careers in, research in accordance with the needs and nature of the research
2. Base research assessment primarily on qualitative evaluation for which **peer review** is central, supported by responsible use of quantitative **indicators**
3. Abandon inappropriate uses in research assessment of **journal- and publication-based metrics**, in particular inappropriate uses of Journal Impact Factor (JIF) and h-index
4. Avoid the use of rankings of research organisations in research assessment

Outline

1. Three contexts for research assessment
2. The purposes of the assessment
3. What should be assessed?
4. The four core commitments
5. What is said about "peer review", "indicators", and "metrics"?
6. The relevance of the Agreement in each context
7. The knowledge base
8. My assessment of the Agreement

Peer review

- Peer review is the most robust method known for assessing quality and has the advantage that it is in the hands of the research community.
- To address the biases and imperfections to which any method is prone, the research community re-assesses and improves peer review practices regularly.
- Moving towards assessment practices that rely more heavily on qualitative methods may require additional efforts from researchers.
- Researchers should be recognised for these efforts and their contributions to reviewing peers' work should be valued as part of their career progression.

Indicators

- Indicators may be used *responsibly*:
 - ... responsible use of quantitative indicators can support assessment where meaningful and relevant, which is context dependent
- Indicators can be used in more qualitative research assessment
 - ... (e.g. narrative and evidence-based CVs, new assessment frameworks and indicators)
- They can even be qualitative:
 - ... qualitative indicators (such as case studies, narratives or statements)
 - ... diversify indicators (Open science badges; Publons, ORCID, open peer review; CRediT; Reporting guidelines e.g. EQUATOR Network) and metrics (Altmetrics, PlumX)
- But no publication-based indicators are mentioned as potentially useful in responsible research assessment

Metrics

- Metrics are mostly used *inappropriately*:
 - ... Inappropriate uses of journal- and publication-based metrics in research assessment should be abandoned (repeated 5 times).
- Metrics *negatively affects* the research culture:
 - ... Assessment processes relying predominantly on journal- and publication-based metrics are known to result in a 'publish or perish' culture that falls short of recognising diverse approaches and could come at the expense of quality
 - ... may negatively affect the quality and impact of research
 - ... may be exacerbated by the pressure on research systems due to the very limited amounts of funding available compared to the pipeline of talented researchers

Metrics = Bibliometrics?

An analysis of core commitment 3

- Not only bibliometrics:
 - ... 'Inappropriate uses' include relying exclusively on author-based metrics (e.g. counting papers, **patents**, citations, **grants**, etc.) to assess quality and/or impact
- But the recurring examples are only from bibliometrics:
 - ...moving away from using metrics like the Journal Impact Factor (JIF), Article Influence Score (AIS) and h-index as proxies for quality and impact.
- Bibliometrics is 'publication-based' by definition:
 - ... Inappropriate uses of **journal- and publication-based metrics** in research assessment should be abandoned.

Bibliometrics is the use of statistical methods to analyse books, articles and other publications, especially in regard with scientific contents. (Wikipedia)

- The words 'statistics' or 'bibliometrics' are never used in the Agreement, only 'publication-based metrics'

Metrics

A comparison with the DORA Declaration and the Leiden Manifesto

- The DORA declaration only addresses the use of **journal-based** bibliometrics in **individual assessment** contexts:
 - ... the need to eliminate the use of journal-based metrics, such as Journal Impact Factors, in funding, appointment, and promotion considerations
- The Leiden Manifesto emphasizes the positive use of bibliometric information in its first principle:
 -Quantitative evaluation should support qualitative, expert assessment. Quantitative metrics can challenge bias tendencies in peer review and facilitate deliberation.
- The Agreement seems to abandon bibliometrics **in general** from research assessment in **all contexts**. There is no mentioning or example of appropriate uses of bibliometrics.

Outline

1. Three contexts for research assessment
2. The purposes of the assessment
3. What should be assessed?
4. The four core commitments
5. What is said about "peer review", "indicators", and "metrics"?
6. The relevance of the Agreement in each context
7. The knowledge base
8. My assessment of the Agreement

The relevance for Recruitment and promotion

Depending on the aims and profile of the research organization, all of these qualifications may be relevant:

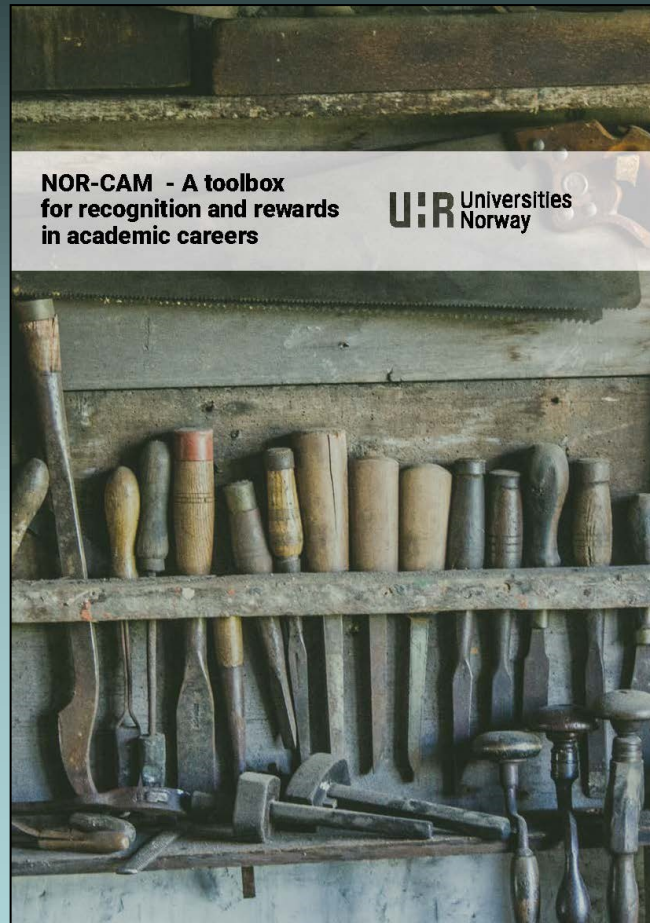
- ...valuable contributions that researchers make to science and for the benefit of society, including diverse outputs beyond journal publications and irrespective of the language in which they are communicated;
- ...practices that contribute to robustness, openness, transparency, and the inclusiveness of research and the research process including: peer review, teamwork and collaboration;
- ...activities including teaching, leadership, supervision, training and mentoring.

The relevance for Recruitment and promotion

Core commitments 1-3 are relevant and needed:

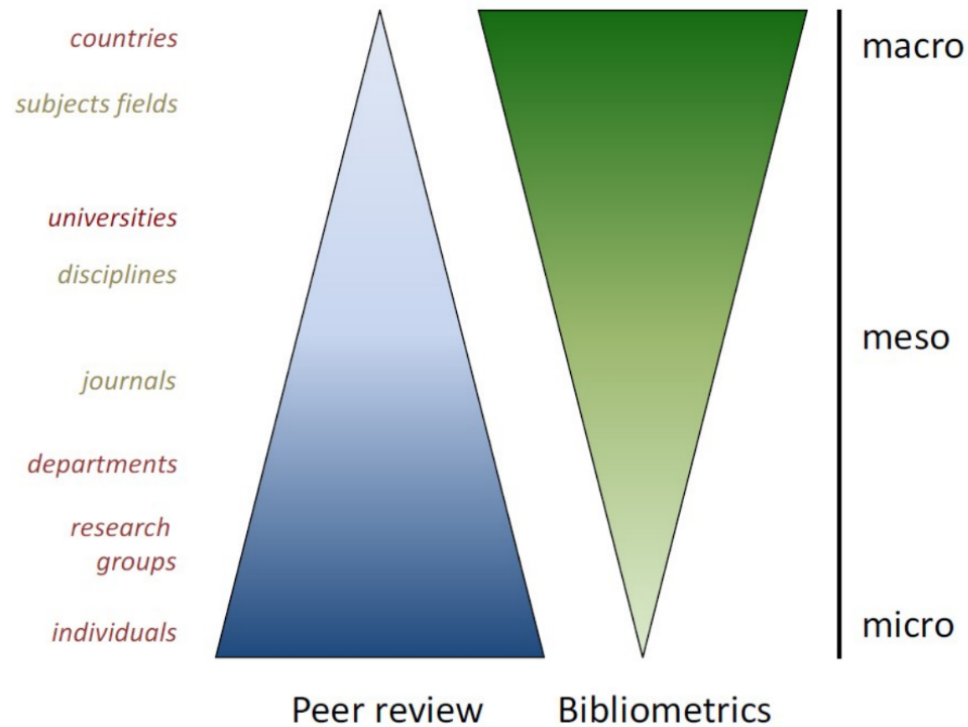
1. Recognise the diversity of contributions to, and careers in, research in accordance with the needs and nature of the research
2. Base research assessment primarily on qualitative evaluation for which peer review is central, supported by responsible use of quantitative indicators
3. Abandon inappropriate uses in research assessment of journal- and publication-based metrics, in particular inappropriate uses of Journal Impact Factor (JIF) and h-index
4. Avoid the use of rankings of research organisations in research assessment

Report available in English at www.uhr.no



Relevance of bibliometrics versus peer judgement

The weight of qualitative (peer evaluation) and quantitative (bibliometrics) methods as function of the aggregation level



Source: GLÄNZEL, 2011

Om bruk av bibliometri på individnivå

Fire råd til forskningsinstitusjoner og forskningsledere fra UHR-Publisering.

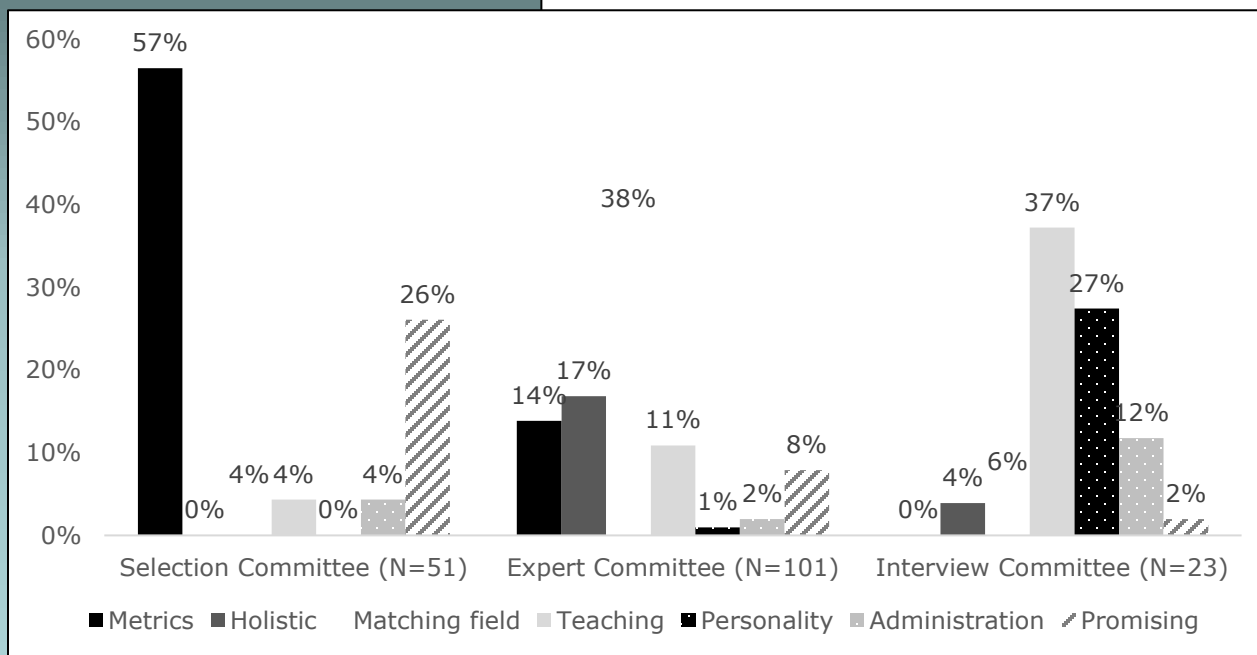
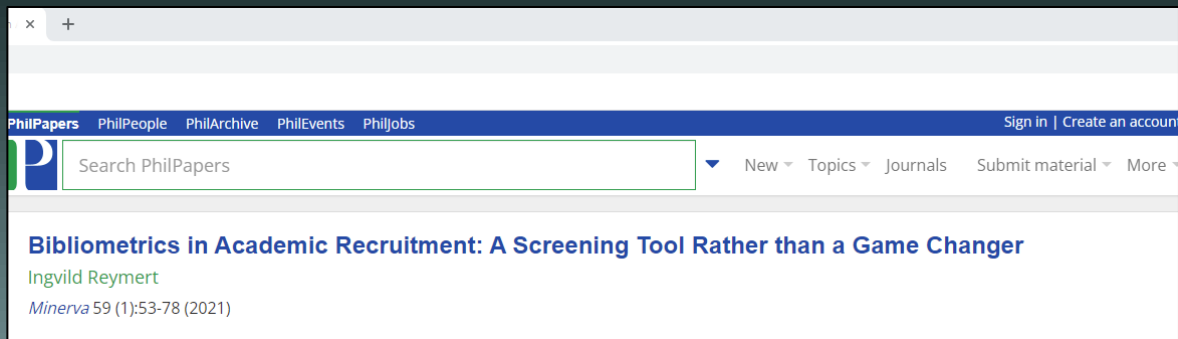
UHR-Publisering, mai 2020

Publiseringsindikatoren er utviklet for å gi oversikt over forskningen på aggregert nivå og er en del av finansieringsmodellen for forskning i universitets- og høyskolesektoren, instituttsektoren og helseforetakene. Den gir årlig statistikk med nasjonal sammenlignbarhet og bidrar med informasjon til evalueringene av norsk forskning.

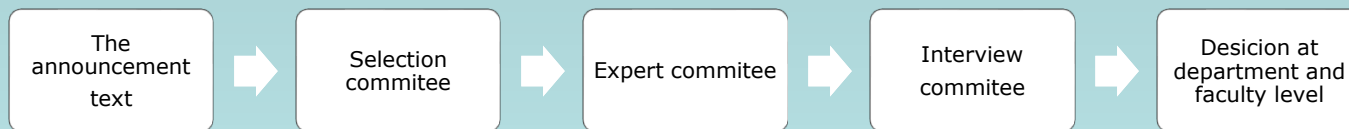
Siden innføringen i 2005 har indikatoren også vært brukt og misbrukt til andre formål lokalt og på individnivå. Evalueringen av indikatoren i 2014 ¹ anbefalte derfor at bruk av indikatoren burde diskuteres mellom forskningsinstitusjonene for å fremme god forskningsledelse og gjensidig læring. UHR-Publisering fulgte opp på flere måter, blant annet ved å publisere et notat med råd om lokal bruk av indikatoren. I mai 2020 har vi revidert notatet, nå med en kort versjon (denne) og en [utvidet versjon](#) som også omtaler Journal Impact Factor (JIF) og H-indeks. For alle tre indikatorer gir vi følgende råd om bruk på individnivå:

- 1. Bibliometri skal ikke brukes alene.** Ansettelse, opprykk, karriereoppfølging og fordelinger av oppgaver og ressurser skal bygge på helhetsvurderinger. I de fleste tilfeller skal ikke bare forskningen vurderes. Men også i vurderinger av forskningen har bibliometriske indikatorer begrenset verdi fordi de kun er tilbakeskuende, ikke tar hensyn til forskjeller i forutsetninger og ikke kan erstatte et beslutningsansvar.
- 2. Bibliometri ser ikke framover.** Bibliometri peker bakover mot tidligere forskningsaktivitet. Vurderinger i forbindelse med ansettelse, opprykk eller tildeling av ressurser skal også se framover og vurdere mulighetene for å innfri utlyste krav og forventninger.
- 3. Bibliometri forstår ikke forskjeller i forutsetninger.** Forskning og faglig aktivitet foregår i mer eller mindre aktive faser alt etter hvilke andre oppgaver man er engasjert i, hvilke ressurser som er tilgjengelige til enhver tid, og hvilken type prosjekter og samarbeid man er med i. Forskningsledere har ansvar for å forstå denne variasjonen ved fordeling av ressurser og ved ansettelse og opprykk.
- 4. Bibliometri kan ikke beslutte.** Når søknadene er mange og det er behov for å foreta et utvalg (screening) før nærmere vurderinger av relevante kvalifikasjoner, kan bibliometri eventuelt være ett av flere hensiktsmessige hjelpemidler. Men verdien av å bruke bibliometri vil være avtakende jo nærmere beslutningen man kommer – se figuren, som viser den avtakende relevansen av kvantitative versus kvalitative metoder i vurdering av forskning på ulike nivåer i forskningssystemet².

Bibliometrics is used responsibly at Norwegian universities



...tial reports with candidate evaluations in economics, sociology, ... explore how metrics were applied in these evaluations in ... most salient criteria in candidate evaluations. Moreover, metrics ... ent for peer review. Contrary to the literature suggesting an ... nce on metrics. In addition, the use of metrics proved strongly ... hese robust evaluation practices provide an empirical example ... rately by isomorphism. Additionally, the disciplinary-dependent ... used but rather translated to fit the local context, resulting in



The relevance for Project funding

In **research** project proposals, the experience and merits documented in **research publications** are most relevant:

- ...valuable contributions that researchers make to science and for the benefit of society, including diverse outputs **beyond journal publications** and irrespective of the language in which they are communicated;

Research publications may also document:

- ...practices that contribute to robustness, openness, transparency, and the inclusiveness of research and the research process including: peer review, teamwork and collaboration;

However, most importantly, the originality and feasibility of the **research proposal itself** is essential in the assessment. This documentation is neglected in the Agreement.

The relevance for Project funding

Only core commitment 2 is directly relevant, at least in Norway and at the EU level, but core commitment 3 may be relevant in other countries:

1. Recognise the diversity of contributions to, and careers in, research in accordance with the needs and nature of the research
2. Base research assessment primarily on qualitative evaluation for which peer review is central, supported by responsible use of quantitative indicators
3. Abandon inappropriate uses in research assessment of journal- and publication-based metrics, in particular inappropriate uses of Journal Impact Factor (JIF) and h-index
4. Avoid the use of rankings of research organisations in research assessment

The relevance for Organizational evaluation

The context of organizational evaluation seems to be included in the Agreement only because of core commitment 4:

4. Avoid the use of rankings of research organisations in research assessment

The primary concern seems to be individual level research assessment:

...Recognising that the international rankings most often referred to by research organisations are currently not 'fair and responsible', the criteria these rankings use should **not trickle down to the evaluation of individual researchers, research teams and research units.**

Individual level effects of the rankings are not possible

The context of organizational evaluation seems to be included in the Agreement only because of core commitment 4:

4. Avoid the use of rankings of research organisations in research assessments

The primary assessment

Recognise research criteria to individual

Scientometrics (2016) 109:2263–2278
DOI 10.1007/s11192-016-2056-5



How can differences in international university rankings be explained?

Fredrik Niclas Piro¹ · Gunnar Sivertsen¹

Received: 19 October 2015 / Published online: 2 July 2016
© Akadémiai Kiadó, Budapest, Hungary 2016

Abstract University rankings are typically presenting their results as league tables with more emphasis on final scores and positions, than on the clarification of why the universities are ranked as they are. Finding out the latter is often not possible, because final scores are based on weighted indicators where raw data and the processing of these are not publically available. In this study we use a sample of Scandinavian universities, explaining what is causing differences between them in the two most influential university rankings: Times Higher Education and the Shanghai-ranking. The results show that differences may be attributed to both small variations on what we believe are not important indicators, as

d to by
the
uation of

The relevance for Organizational evaluation

The description of the purposes of organizational evaluation is mainly focused on summative purposes:

“Allocating funding; public investment accountability; informing decisions on research priorities and improving the definition and implementation of research strategies”

Formative **organizational assessment**

- A professional definition:*
- “An organisational assessment is a systematic process for obtaining valid information about the performance of an **organisation** and the **factors that affect performance**.
- It differs from other types of evaluations because the assessment focuses on **the organisation as the primary unit of analysis**.
- Organisations are constantly trying to adapt, survive, perform, and influence. However, they are not always successful.
- To better **understand what they can or should change to improve their ability to perform**, organisations can conduct organisational assessments.

* As defined by *BetterEvaluation*, a global not-for-profit organisation working to improve the practices worldwide, in the document *Evaluating the Performance of an Organisation* written by Katrina Rojas and Julia Laidlaw.

https://www.betterevaluation.org/en/theme/organisational_performance

The relevance for Organizational evaluation

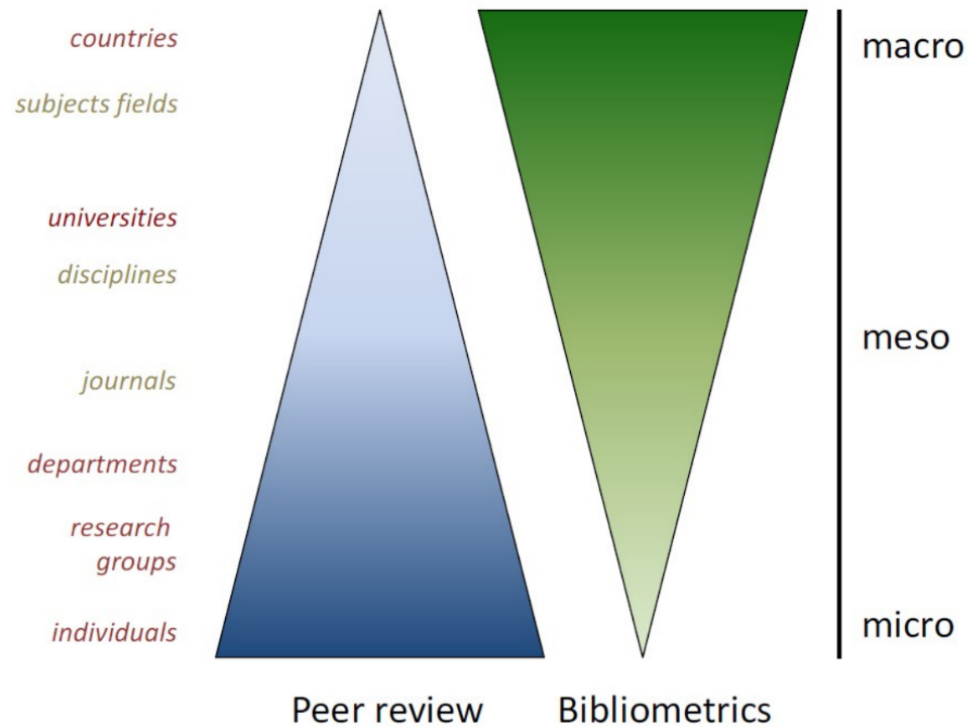
None of the four commitments may actually be relevant:

1. Recognise the diversity of contributions to, and careers in, research in accordance with the needs and nature of the research
2. Base research assessment primarily on qualitative evaluation for which peer review is central, supported by responsible use of quantitative indicators
3. Abandon inappropriate uses in research assessment of journal- and publication-based metrics, in particular inappropriate uses of Journal Impact Factor (JIF) and h-index
4. Avoid the use of rankings of research organisations in research assessment

Instead, professional use of **statistics** and **expertise in organizational evaluation** may be highly relevant.

Relevance of bibliometrics versus peer judgement

The weight of qualitative (peer evaluation) and quantitative (bibliometrics) methods as function of the aggregation level

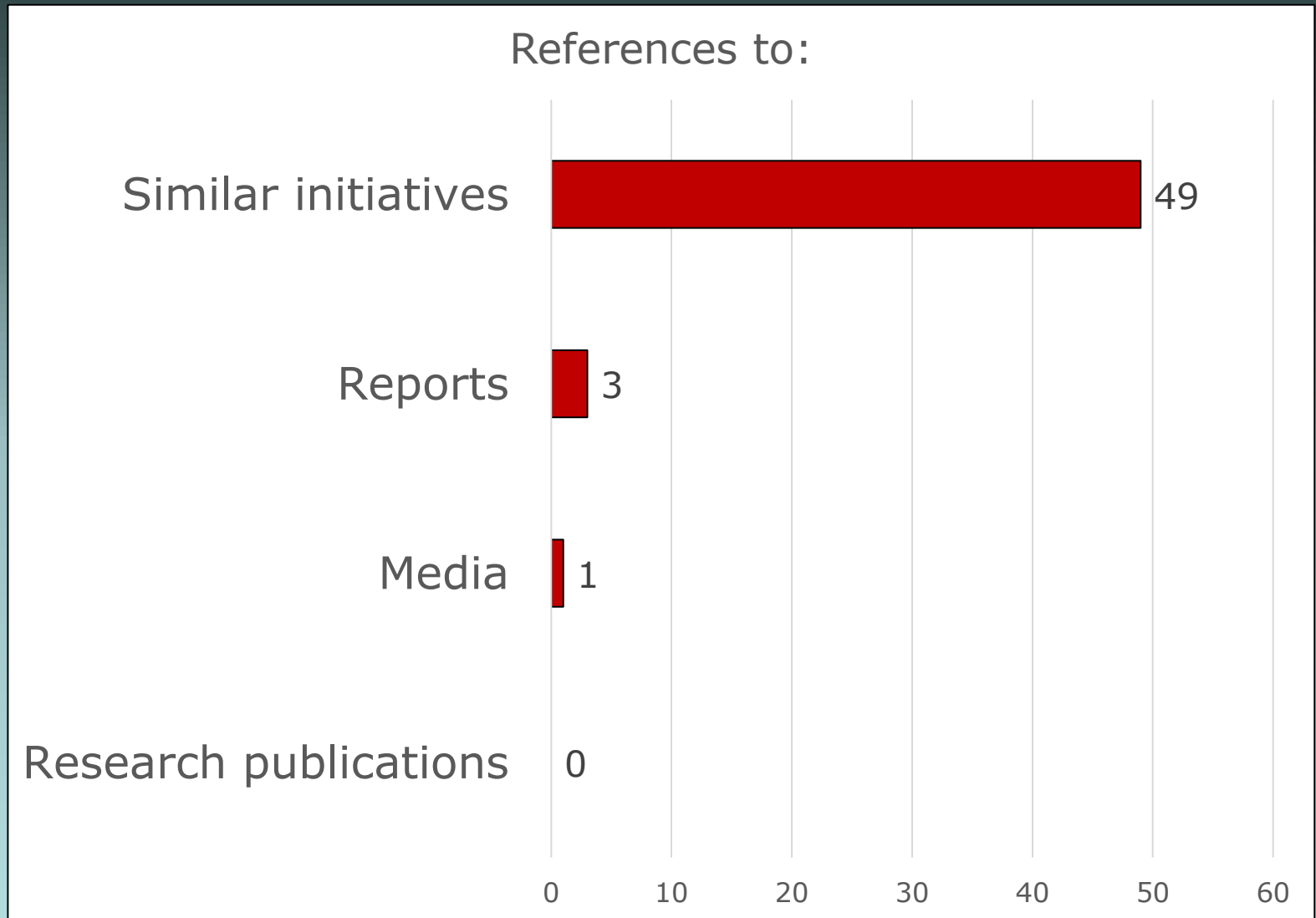


Source: GLÄNZEL, 2011

Outline

1. Three contexts for research assessment
2. The purposes of the assessment
3. What should be assessed?
4. The four core commitments
5. What is said about "peer review", "indicators", and "metrics"?
6. The relevance of the Agreement in each context
7. The knowledge base
8. My assessment of the Agreement

References in the 53 footnotes in the document



No visible contact with relevant research



My assessment of the Agreement

Context	Implemen- tation potential	Under- standing	Relevance and appli- cability	Knowledge base
Organizational evaluation	Weak	Poor	None	Void
Project funding	Strong	Partial	Partial	Anecdotal
Recruitment and promotion	Strong	Good	Good	Anecdotal

The Agreement advocates 'research on research', but only locally within the assessing organizations. There seems to be no contact with the relevant fields of research and their contributions to evidence-based design and appropriate practices and indicators.