



# Informetric methods for studying diversity of the scientific workforce

Towards a State-of-the-Art

Nicolas Robinson-Garcia  
Carmen Corona-Sobrino  
Zaida Chinchilla-Rodríguez  
Daniel Torres-Salinas  
Rodrigo Costas



# Background and rationale

## ★ Genuine interest since the conception of scientometrics on the quantitative study of researchers' activity

- Productivity inequality - Lotka, 1927
- Social stratification - Cole & Cole, 1973
- Hiring & Promotion - Gingras, 2020

## ★ Cautionary remarks for individual metrics in evaluation

- Partial indicators - Martin & Irvine, 1983
- Unit of analysis - Glänzel & Wouters, 2013
- Conceptual flaws - Waltman & van Eck, 2012
- Coverage - van Leeuwen et al, 2001



interpretation of bibliometric results. However, most of these problems can be overcome. **When used properly, bibliometric indicators can provide a "monitoring device" for university research-management and science policy.** They enable research policy-makers to ask relevant questions of researchers on their

### CORRESPONDENCE

#### Impact factors can mislead

Sci. — Impact factors (IFs) for scientific journals, developed by the Institute for Scientific Information (ISI) and published in the section "Journals per category, ranked by Impact Factor" of the *Journal Citation Reports (JCR)*, are frequently used to evaluate the status of scientific journals or even the publication output of scientists. The IF of a journal in year  $T$  is defined as

purchase from ISI. In each category we compared the ranking of journals by IF as printed in the *JCR* to the one based on our correct IF, by calculating the number of journals moving at least 1, 5 or 10 positions. The table shows the first categories affected most severely, measured through the percentage of journals moving at least one position in the ranking. The researchers

#### Individual-level evaluative bibliometrics – the politics of its use and abuse

Wolfgang Glänzel wolfgang.glaenzel@leidenuniv.nl  
Johannes Wouters johannes.wouters@leidenuniv.nl  
Imanet Bakker imab@leidenuniv.nl  
Paul Wouters paul.wouters@leidenuniv.nl

Jointly published by Elsevier Academic Publishers, Dordrecht ScienceMetrics, Vol. 31, No. 1 (2001) 215-248  
and Academic Knowledge Publishers

#### Language biases in the coverage of the Science Citation Index and its consequences for international comparisons of national research performance

THIRD N. VAN LEEUWEN, HENK F. MOED, ROBERT J. W. TISSSEN, MARTIN S. VISSER, ANTHONY F. J. VAN RAAN



THE DOS AND DON'TS IN INDIVIDUAL-LEVEL BIBLIOMETRICS

WOLFGANG GLÄNZEL<sup>1</sup>, PAUL WOUTERS<sup>2</sup>

<sup>1</sup>Centre for B&D Monitoring and Dept. MIS, KU Leuven, Belgium  
<sup>2</sup>Centre for Science and Technology Studies, Leiden University, The Netherlands

The Leiden Manifesto  
for research metrics

ECOOM

KU LEUVEN

# Background and rationale

★ **Diversity, inclusion** and **openness** are the new buzzwords in science policy

- Diversity of outputs and outcomes
- Diversity of knowledges
- Diversity of people/researchers

- Acknowledge and valorise the **diversity in research roles and careers**, including roles outside academia. Value the skills (including open science skills), competences and merits of individual researchers, but also recognise team science and collaboration.
- Ensure **gender equality, equal opportunities and inclusiveness**. Consider gender balance, the gender dimension, and take into account diversity in the broader sense (e.g. racial or ethnic origin, sexual orientation, socio-economic, disability) in research teams at all levels, and in the content of research and innovation.



**I believe in a research culture that recognises a diversity of contributions to science and society; that celebrates high quality and impactful research; and that values sharing, collaboration, integrity and engagement with society, transmitting knowledge from generation to generation.**

**Mariya Gabriel**

Commissioner for Innovation, Research, Culture, Education and Youth



# Background and rationale

## ★ Diversity, inclusion and openness are the new buzzwords in science policy

- Diversity of outputs and outcomes
- Diversity of knowledges
- Diversity of people/researchers

## ★ There is growing sensitivity towards diversity and inequalities in science

- Introduction of new metrics
- New studies quantifying inequalities

Measuring Inequality – Creating an indicator to assess gender bias in universities

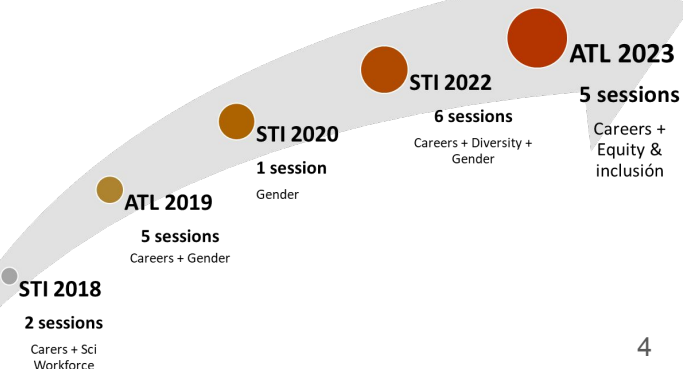
**Indicators**

Type of indicators:  ?

Indicators:  ?

Order by:

- Gender
- Scientific impact
- Collaboration
- Open access
- Gender
- A(M-F)



# Scope of the review

- ★ Many reviews on individual level indicators, but **focused on performance rather than context**
- ★ The raise of **big data** and **machine learning algorithms** provides a great opportunity to study the dynamics of science
- ★ The launch of **new data sources** provides new perspectives and ways into which science can be studied and understood

**What information can metrics provide to study individuals' diversity in science?**

# Scope of the review

## CRITERIA

- ★ Individual level metrics
  - ★ Contextualize rather than assess - *No rankings or league tables!*
- 

## ASPECTS CONSIDERED

- ★ Data sources
- ★ Personal characteristics
- ★ Context
- ★ Team dynamics

# Data sources

Technological and technical developments have been key to the expansion of individual level metrics. We highlight 4 main developments:

## 1. Author name identification

- Databases' author profiles - Scopus AuthorID, WoS Researcher Profiles, Google Scholar Profiles, Dimensions IDs, CWTS disambiguation algorithm
- Author registries - ORCID, CRIS Author profiles

## 2. Metadata quality

- Linkages between fields
- Interoperability between journals and author IDs

## 3. Introduction of author-level features

- Funding
- Academic status
- Contribution statements
- Affiliation history

## 4. Gender / ethnicity identification algorithms

# Individual characteristics

<b>Career stage</b>	<ul style="list-style-type: none"><li>• 1<sup>st</sup> publication year a <i>de facto</i> standard to calculate academic age</li><li>• PhD Year as an alternative when available to calculate academic age</li><li>• Differences by fields noted, also career gaps not considered</li></ul>
<b>Gender</b>	<ul style="list-style-type: none"><li>• Binary approaches based on the use of gender identification algorithms with many approaches used here</li><li>• Attempts made at measuring gender inequality at the publication level</li></ul>
<b>Ethnicity/nationality</b>	<ul style="list-style-type: none"><li>• Most studies use affiliation data to identify nationality or academic origin, mainly used in mobility studies</li><li>• Surnames used to infer ethnicity, less literature in this area in Europe</li></ul>
<b>Cultural identity</b>	<ul style="list-style-type: none"><li>• Area less developed</li><li>• Studies focus on language and country of origin</li><li>• Publication and collaboration patterns by fields (esp. in SSH)</li></ul>



# Context

<b>Trajectory and career</b>	<ul style="list-style-type: none"><li>● Indicators on geographic mobility types of scholars</li><li>● Multiple affiliations with private sector</li><li>● Transitions between sectors</li></ul>
<b>Open practices</b>	<ul style="list-style-type: none"><li>● Open Access</li><li>● Data sharing</li><li>● Open code/software</li></ul>
<b>Funding</b>	<ul style="list-style-type: none"><li>● Studies using funding information at the individual level from:<ul style="list-style-type: none"><li>○ ORCID</li><li>○ Official data sources</li></ul></li></ul>
<b>Societal relevance</b>	<ul style="list-style-type: none"><li>● Altmetrics - Discussions in social media, policy, etc.</li><li>● Inclusion of non-academic outputs</li></ul>

# Team dynamics

<b>Author order</b>	<ul style="list-style-type: none"><li>• 1<sup>st</sup> and last positions as <i>proxies</i> for scientific leadership</li><li>• Papers using corresponding authors, but units of analysis <math>\neq</math> individual</li><li>• Studies on middle authors as <i>proxy</i> for technical roles</li></ul>
<b>Collaboration</b>	<ul style="list-style-type: none"><li>• Measured mainly through co-authorship</li><li>• Indicators characterizing researchers based on the collaboration practices (e.g., churning of co-authors, academic age)</li></ul>
<b>Disciplinary composition</b>	<ul style="list-style-type: none"><li>• Studies measuring team interdisciplinarity as proxy to context or IDR outputs</li><li>• Number of authors influenced by field <math>\leftarrow</math> Team Science</li></ul>
<b>Contributions</b>	<ul style="list-style-type: none"><li>• Types of researchers based on contribution statements</li><li>• Number of contributions by author</li><li>• Changing roles in international collaboration (see van Schalkwyk)</li></ul>

# Conclusions

- ★ Computational developments have been key on their expansion
- ★ Unbalance on the aspects under study and sometimes lack of technical questioning (e.g., accuracy and biases in algorithms)
- ★ Metadata quality but also data access is key for the expansion of these metrics

**Next steps** | Towards a research agenda on the study of diversity of the scientific workforce



# Thank you for your attention

Questions, comments?

Project PID2020-117007RA-I00 funded by:



More on the COMPARE project at

<https://compare-project.eu>

