



ĐO LƯỜNG CÔNG BỐ KHOA HỌC: BỐI CẢNH THẾ GIỚI VÀ LIÊN HỆ VỚI VIỆT NAM

PGS.TS. Hà Quang Thụy, TS. Võ Đình Hiếu, PGS.TS. Phan Xuân Hiếu,
PGS.TS. Nguyễn Trí Thành, GS. TS. Nguyễn Hữu Đức
**PHÒNG THÍ NGHIỆM KHOA HỌC DỮ LIỆU VÀ CÔNG NGHỆ TRI THỨC
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ, ĐẠI HỌC QUỐC GIA HÀ NỘI**

<http://uet.vnu.edu.vn/~thuyhq/>

Ngày 27 tháng 11 năm 2020



Nội dung

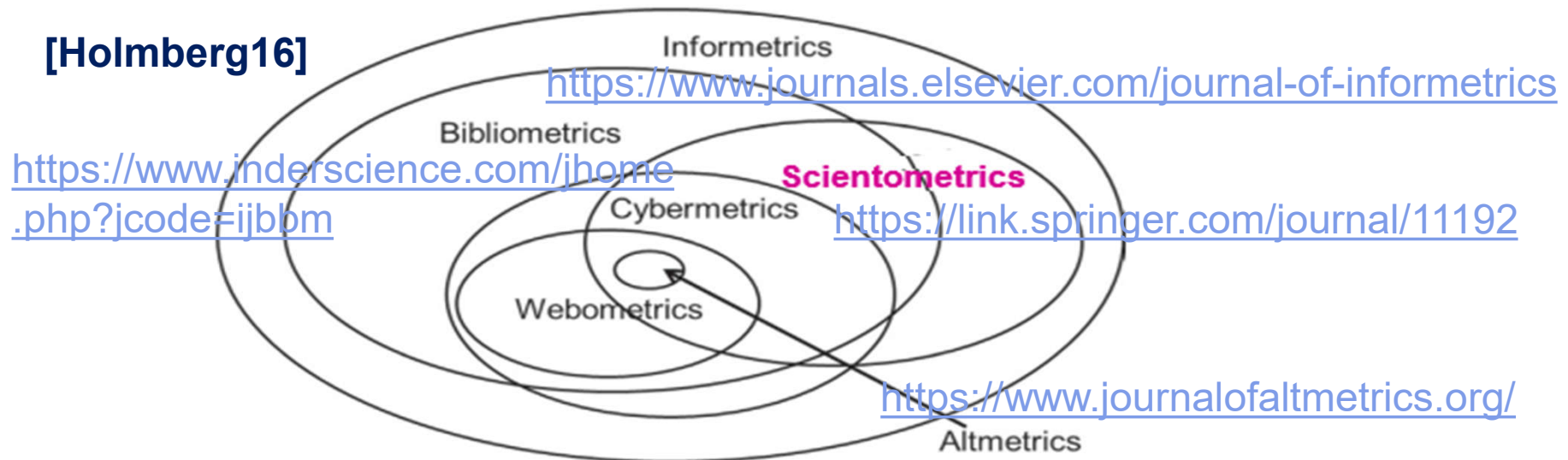
1. Đo lường thông tin và đo lường khoa học
 2. Một số khuyến nghị
 3. Liên hệ với Việt Nam

Đo lường thông tin và đo lường khoa học

- Đo lường khoa học và các lĩnh vực liên quan
- Các cơ sở dữ liệu đo lường khoa học
- Đo lường ảnh hưởng bài báo: số trích dẫn
- Đo lường ảnh hưởng nhà khoa học
- Đo lường ảnh hưởng tạp chí

Đo lường khoa học và lĩnh vực liên quan

[Holmberg16]



● Một khung nhìn đo lường thông tin

- “Đo lường”: “nghiên cứu thống kê định lượng”
- Scientometrics, Bibliometrics và Cybermetrics thuộc Informetrics
- Bibliometrics: tạo, phổ biến và sử dụng thông tin được mã hóa
- Scientometrics: khoa học như một ngành/hoạt động kinh tế.
- Cybermetrics: các tài nguyên, cấu trúc và CNTT trên toàn bộ Internet theo hướng tiếp cận bibliometric và informetric
- Webometrics: Cybermetrics dựa trên Web
- Altmetrics: $\text{Webometrics} \wedge \text{Scientometrics}$

[Holmberg16] Kim Johan Holmberg. *Altmetrics for information professionals: past, present and future*. Chandos Publishing, 2016.

Cơ sở dữ liệu đo lường khoa học

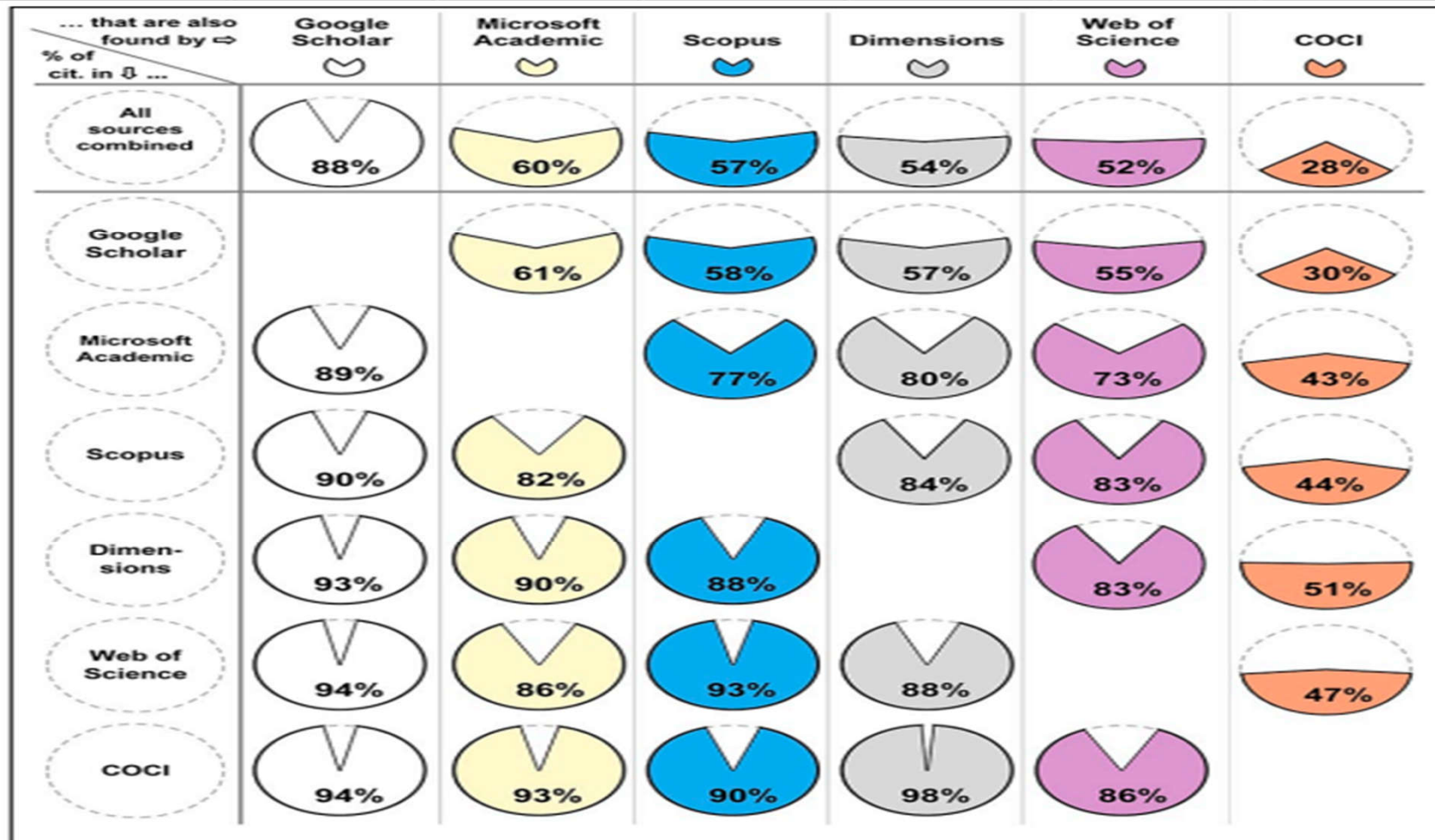
- “Web of Science Core Collection” **WoS** (Institute for Scientific Information: ISI, Clarivate Analytics: <https://clarivate.com/>): Science Citation Index (SCI) 1964, Social Sciences Citation Index 1973, Arts & Humanities Citation Index 1978; đăng ký và chọn lọc. **Thiên vị Bắc Mỹ ?**
- Elsevier’s **Scopus** (11/ 2004): tiếp cận đăng ký-chọn lọc (có loại bỏ). <https://www.elsevier.com/solutions/scopus>. **Thiên vị Châu Âu ?**
- **Google Scholar** (2004, sau Scopus tuần): tự động chỉ mục tài liệu “học thuật” (Google Search), miễn phí, <https://scholar.google.com/>
- Microsoft Academic Search (2006, gỡ bỏ 2012 → 2016 thay bằng nền tảng **Microsoft Academic** (Bing Search), miễn phí như GS, cấp API truy cập hàng loạt, <https://academic.microsoft.com/home>
- **Dimensions** (2018), Digital Science (London), miễn phí chức năng cơ bản, <https://www.digital-science.com/products/dimensions/>
- **COCI** (the OpenCitations Index of Crossref open DOI-to-DOI citations, 2018) do OpenCitations, <https://opencitations.net/index/coci>
- CiteSeerX, ResearchGate, Lens.org, Semantic Scholar, v.v.

[Martín20] Alberto Martín-Martín, Mike Thelwall, Enrique Orduña-Malea, Emilio Delgado López-Cózar. *Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations’ COCI: a multidisciplinary comparison of coverage via citations*. CoRR abs/2004.14329, 2020.

Độ phủ trích dẫn giữa các CSDL [Martín20]

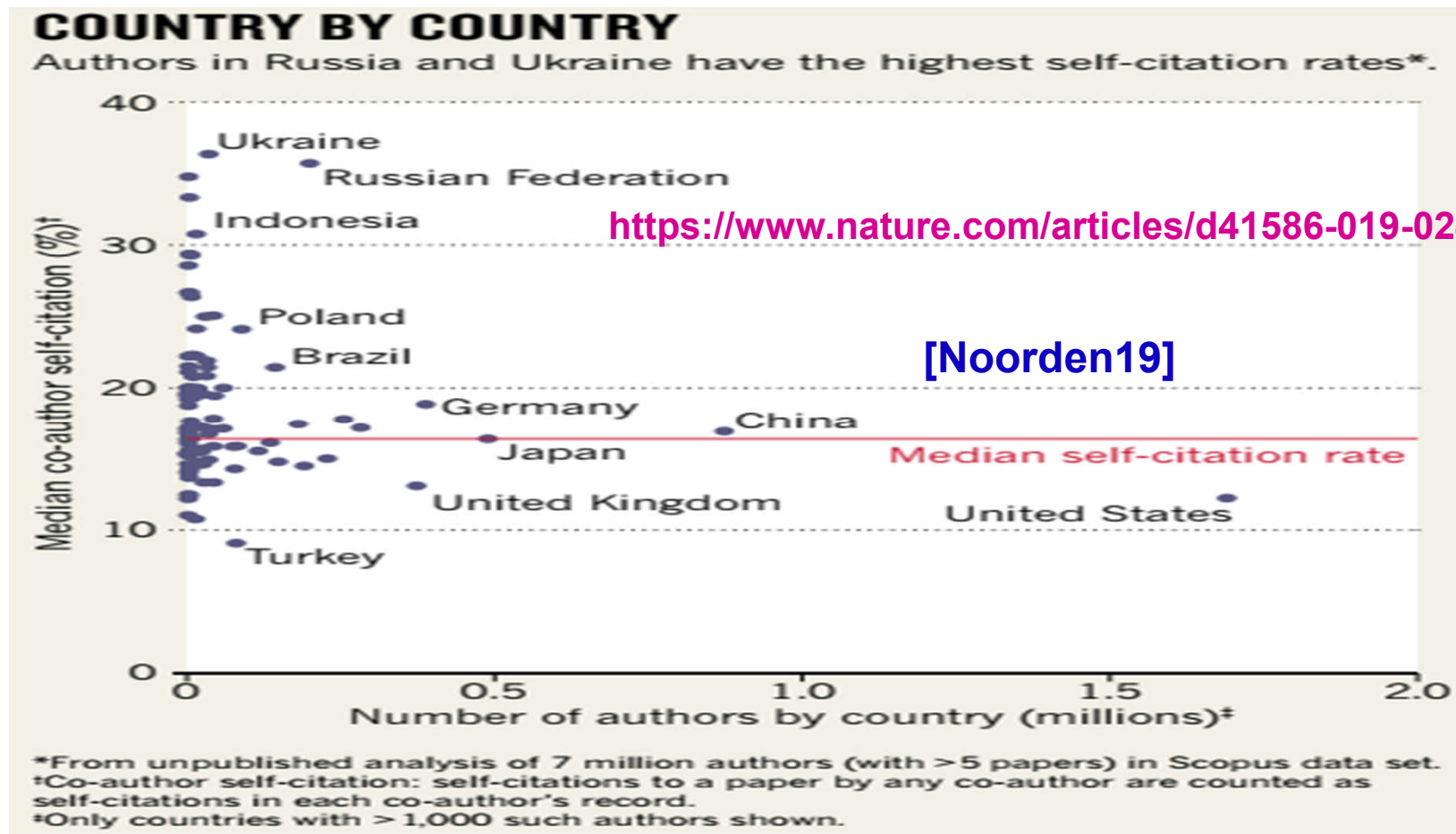
- Khảo sát 3.073.351 chỉ dẫn tới 2.515 tài liệu công bố 2006 được trích dẫn cao đối với 6 CSDL đo lường khoa học

<i>Số TL mẫu</i>	2.515	2.500	2.447	2.478	2.342	2.471
<i>Số chỉ dẫn</i>	2.689.809	1.840.702	1.738.573	1.649.162	1.503.657	852.413



Đo lường ảnh hưởng bài báo: theo số trích dẫn

- Số lượng bài báo khác thuộc một *CSDL đo lường khoa học* tham chiếu tới bài báo (“ảnh hưởng” tới các bài báo khác trong CSDL đó).
- Ngăn ngừa hiện tượng “tự tham chiếu” (self-citing) [Noorden19]



[Noorden19] Richard Van Noorden, Dalmeet Singh Chawla. *Hundreds of extreme self-citing scientists revealed in new database*. Nature 572, 578-579, 2019.

Đo lường ảnh hưởng nhà khoa học: Họ H-index

● Chỉ số H-index

- h-index: h là **số nguyên lớn nhất** mà có h bài báo của nhà khoa học được ít nhất h bài báo khác trích dẫn.
- h-index “cân đối” số bài báo và ảnh hưởng bài báo; cải tiến các độ đo sử dụng: *tổng số bài báo*, *tổng số bài tham chiếu*, *số tham chiếu mỗi bài báo*, *số bài báo có ý nghĩa*, v.v.

● Họ chỉ số H-index

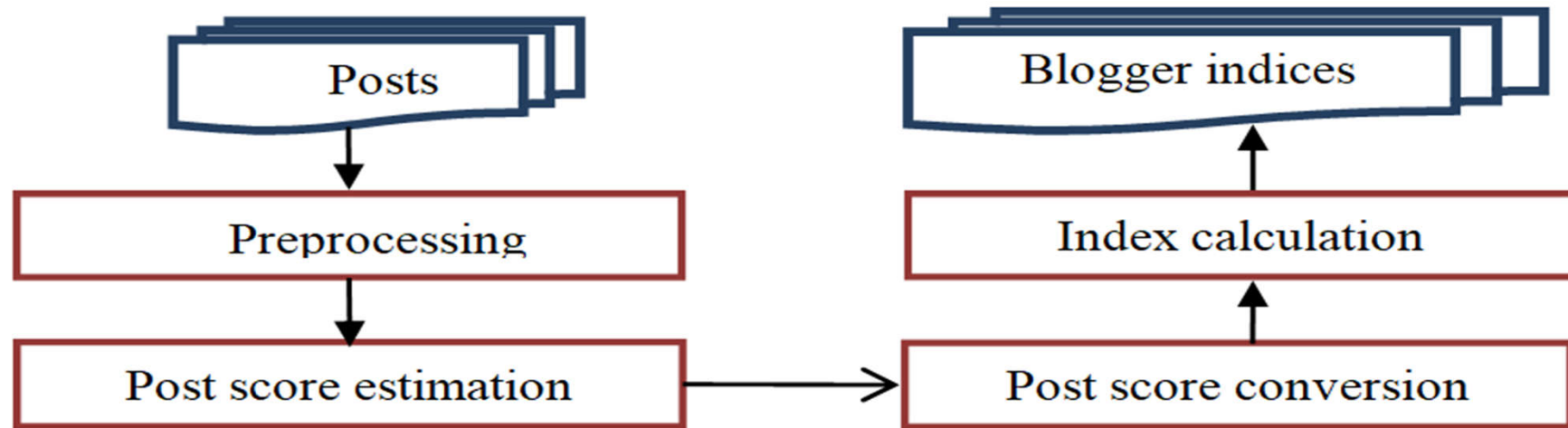
- h-index có một vài hạn chế (bỏ qua ảnh hưởng của h bài báo ảnh hưởng nhất, độ tươi “ưu tiên cái cũ”, v.v.) → cần cải tiến.
- g-index: tổng số tham chiếu của tập g bài báo $\geq g^2$ (có $g \geq h$). Aminer <https://www.aminer.org/> chỉ dẫn cả h-index và g-index.

- $a\text{-index} = \frac{1}{h} \sum_{j=1}^h cit_j$; $R\text{-index} = \sqrt{\sum_{j=1}^h cit_j}$; $\pi\text{-index}$

● Một vài độ đo của Google Scholar

- i10 (i20): số lượng bài báo có ít nhất 10 (20) tham chiếu.
- I10, i20 trong năm năm gần nhất: “độ tươi” ảnh hưởng của nhà KH.

Họ H-index: áp dụng tính toán mạng xã hội



● Xác định blogger ảnh hưởng nhất

- Blogger có ảnh hưởng $k \leftrightarrow$ h-index nhà KH là k (bài báo \leftrightarrow bài viết)
- Sự thừa nhận (inlinks): lượng bài viết khác tham chiếu đến
- Độ phát sinh hoạt động (comments): lượng bình luận nhận được
- Tính mới (outlinks): lượng liên kết ngoài từ bài viết
- Độ thuyết phục: độ dài, phong cách viết, sử dụng từ ngữ, v.v.
- **Input:** M blogger $\{b_1, b_2, \dots, b_M\}$, tập L bài viết của M blogger đó $\{p_1, p_2, \dots, p_L\}$; **Output:** Tập K blogger có điểm ảnh hưởng cao nhất.

[Bui14] Dinh-Luyen Bui, Tri-Thanh Nguyen, Quang-Thuy Ha. *Measuring the Influence of Bloggers in Their Community Based on the H-index Family*. ICCSAMA 2014: 313-324. (10 Scopus citations, trong đó có Xun Liang. *Social Computing with Artificial Intelligence*. Springer 2020)

Đo lường ảnh hưởng tạp chí

● Hệ số ảnh hưởng tạp chí

- Journal Impact Factors (JIF): Số lượng trích dẫn trung bình trong năm được tính (y) của các bài báo được công bố trên tạp chí đó ở hai năm quá khứ kề cận (y-1, y-2).
$$JIF_y = \frac{Citation_y}{Publications_{y-1} + Publications_{y-2}}$$

● Ảnh hưởng bài báo theo ảnh hưởng tạp chí

- Về chính sách đầu tư công bố: ưu tiên đăng tạp chí JIF cao, áp lực gửi kết quả nghiên cứu, tiềm ẩn nguy hiểm “công bố trên các tạp chí cụ thể” dẫn tới thiên vị tạp chí WoS hoặc tạp chí tiếng Anh. Liên hợp quốc, Ngân hàng thế giới, Diễn đàn kinh tế thế giới **sử dụng Scopus**.
- Về đánh giá bài báo hoặc cá nhân nhà khoa học: Anthony van Raan “*sử dụng JIF để đánh giá hiệu năng nghiên cứu đối với một bài báo hoặc một cá nhân - đó là một tội trọng*”.
- JIF có thể là “bí mật thương mại” (<http://altmetrics.org/manifesto/>), đã trở thành “thương hiệu”, một hàng hóa “xa xỉ” → “hàng nhái”.
- JIF: nguồn thông tin giá trị cao song không thể được sử dụng bừa bãi.

[Larivière19] Vincent Larivière, Cassidy R. Sugimoto. *The Journal Impact Factor: A Brief History, Critique, and Discussion of Adverse Effects*. In (Wolfgang Glänzel, Henk F. Moed, Ulrich Schmoch, Mike Thelwall. *Springer Handbook of Science and Technology Indicators*. Springer, 2019), pp. 3-24.

Một số khuyến nghị quốc tế

- Tuyên bố DORA
- Khuyến nghị của Ủy ban IEEE
- Tuyên ngôn Leiden

Tuyên bố DORA

● Giới thiệu và ba chủ đề chính

- Declaration On Research Assessment, Tuyên bố San Francisco 2012.
- Cần **loại bỏ việc sử dụng những độ đo dựa trên tạp chí**.
- Cần đánh giá nghiên cứu khoa học dựa trên **phẩm chất của chính nó**.
- Cần tận dụng các cơ hội do **xuất bản trực tuyến** tạo ra và khám phá **các độ đo mới**.

● Mười tám khuyến nghị

- Khuyến nghị chung: *Không dùng các độ đo đánh giá tạp chí (ví dụ JIF), như độ đo thay thế cho việc đánh giá đối với chất lượng của một bài báo khoa học, để từ đó đánh giá sự đóng góp của một nhà khoa học, hoặc để xem xét việc tuyển dụng, thăng tiến, hay tài trợ.*
- Khuyến nghị nhà tài trợ (2): (i) tiêu chí đánh giá hiệu năng, (ii) hình thức kết quả nghiên cứu và nhiều độ đo ảnh hưởng đa dạng
- Khuyến nghị tổ chức KH (2): tiêu chí rõ ràng, nhiều độ đo ảnh hưởng
- Khuyến nghị nhà xuất bản (5), Khuyến nghị tổ chức đo lường khoa học (4), Khuyến nghị cá nhân nhà KH (4)

Khuyến nghị của Ủy ban IEEE

● Giới thiệu và các luận điểm

- JIF là một cấu trúc phức tạp, đa chiều → cần có tổ hợp nhiều độ đo
- Chỉ nên áp dụng **nhóm tập thể** (không riêng lẻ) **chỉ số hiệu năng KH** kết hợp với **đánh giá đồng cấp** theo **quy tắc ứng xử rõ ràng**.
- Tính đặc thù theo lĩnh vực

● Khuyến nghị với Công nghệ, KHMT và CNTT

- Dùng nhiều độ đo bổ sung là quan trọng để cung cấp cái nhìn phù hợp, toàn diện và cân bằng. Ngoài JIF, IEEE cung cấp Eigenfactor và Article Influence; hoan nghênh các độ đo Altmetrics.
- Mọi độ đo dựa trên tạp chí đã không được thiết kế để nắm bắt chất lượng bài báo riêng lẻ → không được dùng để đại diện chất lượng bài báo và do đó đánh giá cá nhân nhà KH.
- Dù độ đo KH có thể được dùng như nguồn thông tin bổ sung để đánh giá chất lượng trong một lĩnh vực NC cụ thể, **cách thức chính đánh giá** một dự án NC hoặc cá nhân nhà KH phải là **đánh giá đồng cấp** dựa trên việc **xem xét nội dung KH là khía cạnh quan trọng nhất**, là kỳ vọng xuất bản theo lĩnh vực, quy mô và cộng đồng.

[IEEE14] IEEE. *Appropriate use of bibliometric indicators for the assessment of journals, research proposals, and individuals*. IEEE Computer Graphics and Applications, Vol.34, Issue 2, Mar.-Apr. 2014.

Tuyên ngôn Leiden

● Mười nguyên tắc

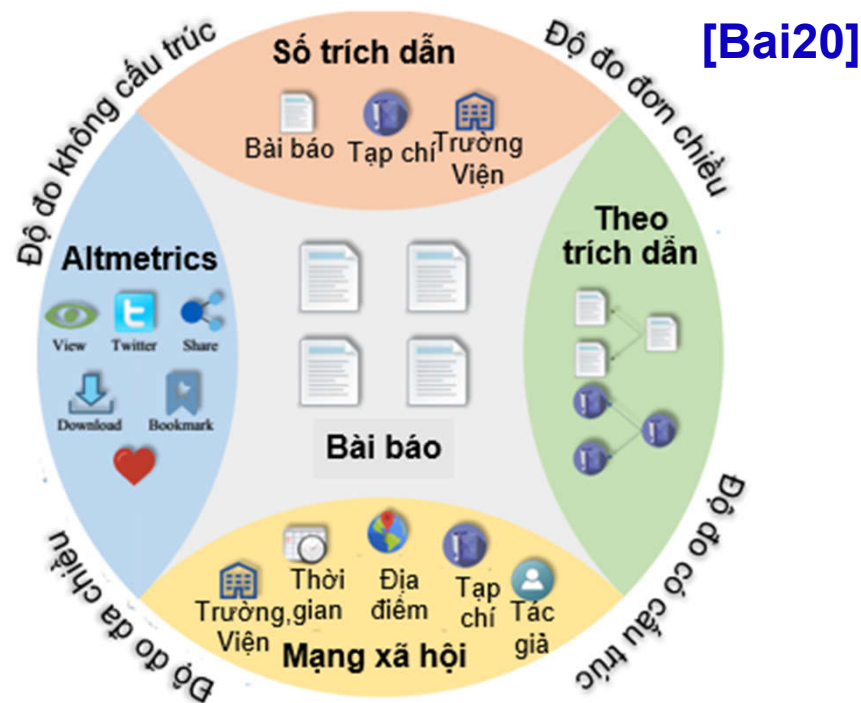
- Đánh giá định lượng nên hỗ trợ đánh giá định tính, chuyên gia.
- Đo lường hiệu năng phù hợp với các nhiệm vụ nghiên cứu của tổ chức, nhóm hoặc nhà nghiên cứu.
- Ưu tiên độ xuất sắc đối với các nghiên cứu liên quan tại địa phương.
- Đảm bảo quá trình thu thập và phân tích dữ liệu công khai, minh bạch và đơn giản.
- Cho phép người được đánh giá tự xác minh và phân tích dữ liệu.
- Cần diễn giải sự thay đổi theo lĩnh vực trong thực tiễn xuất bản và trích dẫn. Ví dụ, JIF hàng đầu toán học là 3, sinh học tế bào là 30.
- Cơ sở đánh giá cá nhân nhà KH theo định tính hồ sơ của họ.
- Tránh đặt sai các tính toán cụ thể và độ chính xác sai.
- Nhận biết tác động hệ thống của việc đánh giá và các chỉ số đo lường.
- Thường xuyên xem xét kỹ lưỡng các chỉ số đo lường và cập nhật chúng.

[Hicks15] Diana Hicks, Paul Wouters, Ludo Waltman, Sarah de Rijcke, Ismael Rafols. *Bibliometrics: The Leiden Manifesto for research metrics*. Nature, Volume 520, Issue 7548, Article 22 April 2015.

Định hướng đo lường ảnh hưởng bài báo



<http://altmetrics.org/manifesto/>



- **Tuyên ngôn Altmetrics**

- Bốn thành phần
- Ấn phẩm, bài báo có tải và xem.

- **Một khung nhìn đa thành phần [Bai20]**

- Hai cặp bộ độ đo cấu trúc – không cấu trúc, đơn chiều – đa chiều.
- Đa tiêu chí: Số trích dẫn, theo trích dẫn, mạng xã hội, Altmetrics

[Bai20] Xiaomei Bai, Hui Liu, Fuli Zhang, Zhaolong Ning, Xiangjie Kong, Ivan Lee, Feng Xia. *An Overview on Evaluating and Predicting Scholarly Article Impact*. CoRR abs/2008.03867, 2020.

Liên hệ với Việt Nam

- Một vài quan sát
- Nhóm nghiên cứu và hệ thống mở
- Quản lý nhà nước và tổ chức KH

Một vài quan sát

● Trao đổi về tuyên bố DORA

- Nguyễn Đức Thế [The13] có bài trao đổi đầu tiên về DORA, có đề cập **hiện tượng lạm dụng JIF tại Việt Nam** và hạn chế từ chính sách của Nafosted khuyến khích công bố trên ISI.
- Vũ Cao Đàm [Dam14] bàn luận nhân bài viết của Nguyễn Đức Thế, đề cập tới thói quen “**cứ nói đến đánh giá nghiên cứu khoa học là nói đến chỉ số IF của ISI**” và tâm lý sợ trao đổi về thói quen này.

● Các quan sát khác

- Các sự kiện xét công nhận chức danh giáo sư, luận án Tiến sỹ: xuất hiện hiện tượng “đại chúng hóa” đánh giá khoa học.
- Quan niệm không đúng về công bố tạp chí mở. Cần quan niệm “tạp chí mở” tương ứng với **minh bạch, cho phép áp dụng Altmetrics và các độ đo đa dạng [Bai20] như các khuyến nghị quốc tế**.
- Quan niệm định kiến gắn chất lượng bài báo với lệ phí công bố bài báo, nhà xuất bản, v.v.

[The13] Ngô Đức Thế. *Tuyên bố DORA có làm thay đổi cách đánh giá khoa học?* Tạp chí Khoa học Công nghệ Việt Nam, 12(2013): 31-34.

[Dam14] Vũ Cao Đàm. *Bình luận nhân bài viết của Ngô Đức Thế “Tuyên bố DORA có làm thay đổi cách đánh giá khoa học?”*. Trang web Hội đồng chức danh giáo sư nhà nước, truy nhập ngày 20/5/2017

Nhóm nghiên cứu và hệ thống mở

● Nhóm nghiên cứu

- Hoạt động KH-CN gắn với **nhóm nghiên cứu** (NNC), ngoại trừ một vài chuyên ngành lý thuyết hẹp.
- NNC = Nghiên cứu + đào tạo sau, trên đại học **chất lượng cao**
- Ảnh hưởng nhà KH = ảnh hưởng của NNC + ảnh hưởng của nhà KH đối với NNC (là chủ trì, thành viên chủ trì, cốt lõi). Ví dụ, nhóm nghiên cứu của GS. TS. Nguyễn Hữu Đức (Scopus h-index là 24), GS. TSKH. Nguyễn Đình Đức (Scopus h-index = 26), GS. TS. Lê Bảo Long (Canada, có NCS + TTS Việt Nam, Scopus h-index = 38).
- Nhóm nghiên cứu → Tiềm năng về trường phái khoa học Việt Nam.
- Tránh hiện tượng nhà KH “có ảnh hưởng công bố lớn” lại thiếu “NNC nội tại” → nghi ngờ nỗ lực nghiên cứu.

● Hệ thống mở công trình khoa học

- Cá nhân nhà KH **công bố mở** công trình KH, chẳng hạn, **định kỳ hiệu chỉnh trang Google Scholar cá nhân** (ví dụ, ĐHQGHN).
- Khuyến khích các tổ chức KH công nghệ **công bố mở** thành tựu KH.
- Hình thành tổ chức **đo lường KH** của Việt Nam, chẳng hạn nâng cấp **thành phần thống kê bài báo** trong **hệ thống xếp hạng đại học**

Quản lý nhà nước và tổ chức KH-CN

● Đầu tư, giám sát, đánh giá tổ chức KH-CN

- Đầu tư KH-CN (0.4% GDP), giáo dục đại học (0.33% GDP) ↑
- Giám sát, đánh giá **tổ chức KH-CN** phục vụ đầu tư hiệu quả
- Hạn chế trực tiếp tổ chức thực hiện CT, dự án, đề tài KH-CN.
- Chuyên nghiệp hoạt động giám sát, đánh giá tổ chức KH-CN
- Chuyển hoạt động **đánh giá cá nhân nhà KH** cho tổ chức KH quản lý nhà KH: chẳng hạn, chức danh giảng dạy đại học (giáo sư, phó giáo sư) về trường đại học và nhà nước phong danh hiệu thành tựu.

● Tổ chức KH-CN

- Trường ĐH định hướng nghiên cứu, tập trung **chiến lược nhóm nghiên cứu nội tại** tiến tới nhóm NC cộng tác khoa học quốc tế
- Đầu tư giai đoạn hiện nay: **khoán sản phẩm cho nhóm NC**
- Trực tiếp đánh giá, bố trí chức danh **nhà KH hướng nhóm NC**: GS chủ trì NNC mạnh, tổ chức tốt ĐT sau-trên ĐH & công bố quốc tế
- Tự giám sát hoạt động KH-CN (nòng cốt là cộng đồng nhà khoa học nội bộ), kết hợp với cơ quan quản lý cấp trên

Một số bài viết liên quan của nhóm tác giả

Những yếu tố quan trọng trong Chiến lược quốc gia về trí tuệ nhân tạo: Tạp chí Khoa học và Công nghệ Việt Nam, tháng 10/2020, <https://vjst.vn/vn/tin-tuc/3952/nhung-yeu-to-quan-trong-trong-chien-luoc-quoc-gia-ve-tri-tue-nhan-tao.aspx>. *Phiên bản mở rộng*: <https://uet.vnu.edu.vn/~thuyhq/PPNCKH/Mot-phiên-ban-gui-Tap-chi-KH&CN.pdf>

Môi trường và giải pháp cho chiến lược quốc gia về trí tuệ nhân tạo của Việt Nam: Tạp chí Khoa học và Công nghệ Việt Nam, tháng 08/2020, <http://vjst.vn/vn/tin-tuc/3283/moi-truong-va-giai-phap-cho-chien-luoc-quoc-gia-ve-tri-tue-nhan-tao-cua-viet-nam.aspx> . *Phiên bản mở rộng*: https://uet.vnu.edu.vn/~thuyhq/PPNCKH/Phien-ban_09-06-2020.pdf.

Đại học thông minh: Bối cảnh thế giới và liên hệ với Việt Nam. Trang web Trường Đại học Công nghệ (ĐHQGHN) <https://uet.vnu.edu.vn/dai-hoc-thong-minh-boi-can-h-the-gioi-va-lien-voi-viet-nam/>. https://uet.vnu.edu.vn/~thuyhq/PPNCKH/DHTM_VNU.pdf

Kinh tế số: Bối cảnh thế giới và liên hệ với Việt Nam: Tạp chí Công thương, tháng 04/2020, <http://www.tapchicongthuong.vn/bai-viet/kinh-te-so-boi-can-h-the-gioi-va-lien-he-voi-viet-nam-70275.htm>

Trí tuệ nhân tạo trong thời đại số: Bối cảnh thế giới và liên hệ với Việt Nam: Tạp chí Công thương, tháng 8/2018, <http://tapchicongthuong.vn/bai-viet/tri-tue-nhan-tao-trong-thoi-dai-so-boi-can-h-the-gioi-va-lien-he-voi-viet-nam-55038.htm>

Khoa học dữ liệu, tiền hóa dữ liệu và khả năng thực thi tại Việt Nam: Tạp chí Công thương-Các kết quả nghiên cứu khoa học và Ứng dụng công nghệ, 3, trang 160-165, 2018, <https://uet.vnu.edu.vn/~thuyhq/PPNCKH/Tap-chi-Cong-thuong-NCKH-so-3-thang-3-2018.pdf>

Xin trân trọng cảm ơn!