

The Future of Scientific Integrity in an Open Science: Challenges and Prospects

Margaux Larre-Perez, Ph.D.
Université Paris Cité Libraries
Université Paris Cité, Paris, France.





Université Paris Cité - in a nutshell



Historical building of Université Paris Descartes medical school currently the University headquarters (where I work) (Source: Université Paris Cité wikipedia page)

- Public research university in Paris, France (technically an "experimental establishment")
- Result of the merger of two universities: Université Paris Descartes (known as Paris 5) and Université Paris Diderot (known as Paris 7)
- More than 62k students
- About 3500 graduate students



What awaits you in this webinar

How open science practices help fostering scientific integrity.

How an increasingly open science brings along new challenges to scientific integrity.

The future of scientific integrity challenges?



Introduction: a few words on Open Science

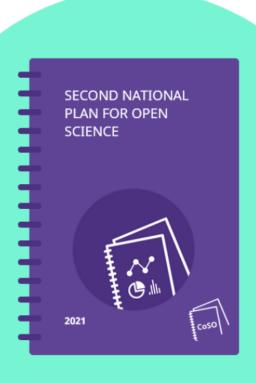




"Open Science is the practice of science in such a way that others can collaborate and contribute, where research data, lab notes and other research processes are freely available, under terms that enable reuse, redistribution and reproduction of the research and its underlying data and methods."

https://www.fosteropenscience.eu/foster-taxonomy/open-science-definition



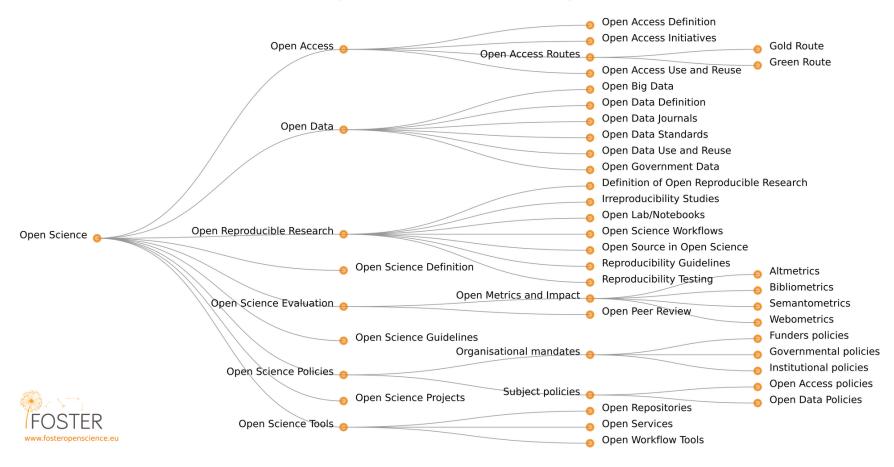


PUBLIC POLICIES

Open science refers to the unhindered dissemination of results, methods and products from scientific research. It draws on the opportunity provided by recent digital progress to develop open access to publications and - as much as possible - data, source code and research methods.



Open Science Taxonomy



- A worldwide movement for knowledge dissemination.
- A plethora of activities in many domains, from publishing to software development through citizen science.
- A beneficial scientific practice for me as a young researcher and for society.

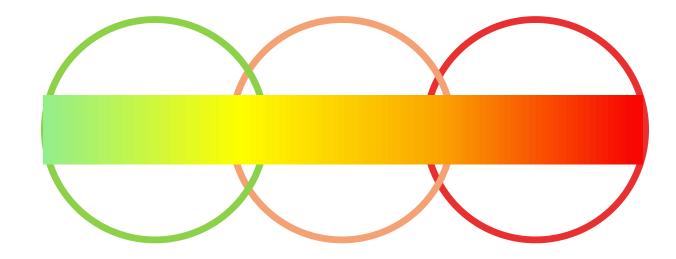


1. Fostering scientific integrity through Open Science.



Scientific Integrity: a spectrum of scientific behaviors

Responsible Research Practices (RRP)



Scientific Misconduct (fraud)

Traditionally includes: plagiarism, falsification and fabrication.

Questionable Research Practices
(QRP) (John et al., 2012): behaviors going from genuine research misknowledge to intentionnally bad practices. Evolving definition to include new or overlooked behaviors.



QRPs as prevalent practices accross countries and disciplines



Questionable research practices in ecology and evolution

Hannah Fraser . Tim Parker, Shinichi Nakagawa, Ashley Barnett, Fiona Fidler

Published: July 16, 2018 • https://doi.org/10.1371/journal.pone.0200303



Both Questionable and Open Research Practices Are Prevalent in Education Research

Matthew C. Makel Makel

Volume 50, Issue 8 | https://doi.org/10.3102/0013189X211001356 | View correction

Communication Methods and Measures >

Volume 9, 2015 - Issue 4: Questionable Research and Publication Practices in Communication Science

Questionable Research Practices in Experimental Communication Research: A Systematic Analysis From 1980 to 2013

Jörg Matthes S, Franziska Marquart, Brigitte Naderer, Florian Arendt, Desirée Schmuck & Karoline Adam Pages 193-207 | Published online: 30 Nov 2015

PLOS ONE

Questionable research practices among italian research psychologists

Franca Agnoli , Jelte M. Wicherts, Coosje L. S. Veldkamp, Paolo Albiero, Roberto Cubelli

Published: March 15, 2017 • https://doi.org/10.1371/journal.pone.0172792



QRPs as prevalent practices accross research roles.



The survey found Ph.D. students had the hardest time meeting the standards of responsible research. Some 53% of them admitted to frequently engaging in one of the 11 questionable research behaviors within the past 3 years, compared to 49% of associate and full professors.



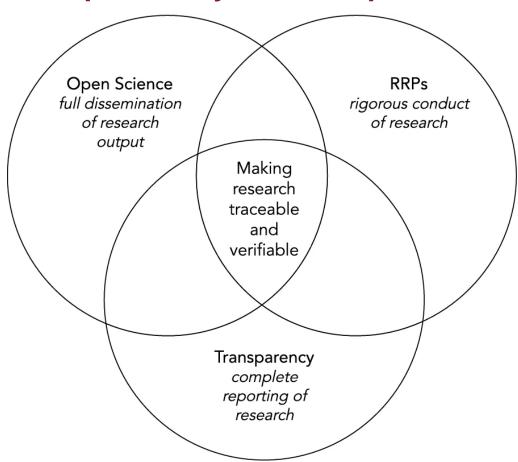
Davide Bonazzi/Salzman Art



Open Science and research integrity: a complementary relationship

"these concepts [research integrity, rigorous research practices, transparency] are all crucial to strengthen trust in research and researchers by making research more traceable and verifiable. We believe that their focus on a particular phase of the research process can be instrumental to further the understanding of these concepts, because it is precisely by virtue of their different foci that these concepts become complementary and mutually reinforcing"

"In a nutshell, we believe that responsible research practices focus more on the rigorous conduct of research, transparency focuses predominantly on the complete reporting of research at every stage of the research lifecycle, and open science's core focus is mostly on the dissemination of research."





The Replication Crisis in psychology

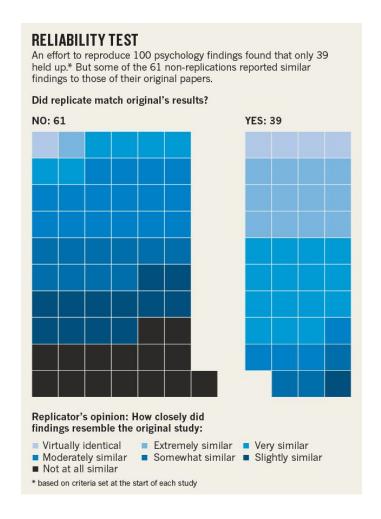
The Atlantic

SCIENCE

How Reliable Are Psychology Studies?

A new study shows that the field suffers from a reproducibility problem, but the extent of the issue is still hard to nail down.

By Ed Yong





The Replication Crisis: not just in psychology





2%

experiments with open data

70%

of experiments required asking for key reagents

69%

of experiments needing a key reagent original authors were willing to share

0%

of protocols completely described

32%

of experiments the original authors were not helpful (or unresponsive)

41%

of experiments the original authors were very helpful



The Replication Crisis: not just in psychology

"But in several papers, researchers failed to properly separate the pools of data used to train and test their code's performance, a mistake termed "data leakage" that results in a system being tested with data it has seen before, like a student taking a test after being provided the answers."

Wired, 10/08/2022 (see next slide)



Context

Many quantitative science fields are adopting the paradigm of predictive modeling using machine learning. We welcome this development. At the same time, as researchers whose interests include the strengths and limits of machine learning, we have concerns about reproducibility and overoptimism.



The Replication Crisis: not just in psychology

nature

NEWS 26 July 2022

Could machine learning fuel a reproducibility crisis in science?

'Data leakage' threatens the reliability of machine-learning use across disciplines, researchers warn.

Elizabeth Gibney



WILL KNIGHT

BUSINESS AUG 10, 2022 7:00 AM

Sloppy Use of Machine Learning Is Causing a 'Reproducibility Crisis' in Science

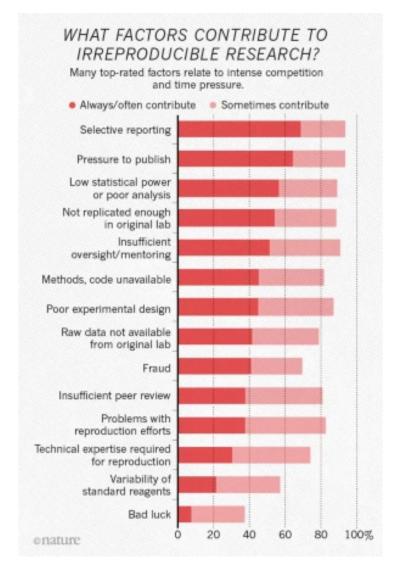
All hype has researchers in fields from medicine to sociology rushing to use techniques that they don't always understand—causing a wave of spurious results.



Causes for the Replication crisis

A few possible explanations:

- **Positive publication biais** (i.e. tendancy of scientific journals to only publish statistically significant results), not encouraging researchers to share negative, non replicative results.
- *Questionable research practices*, espacially data snooping, even fraud.
- *Inability to properly reproduce research artifacts* (e.g. method, stimuli).



https://www.nature.com/articles/533452a



If for any given experiment the rationale, datasets and methods are freely available, then the robustness of scientific results can be constantly reassessed.

Researchers could then:

- **attempt to replicate the original experiment**, i.e. rerun the original method and compare obtained data with the original experiment's.
- **attempt to re-analyse the original data**, i.e. re-run the statistical tests described by the original method, and compare obtained results with the original experiment's
- **avoid "wasting" time** on further exploring a counter-intuitive result which looked promising but actually resulted from statistical chance or a poorly designed procedure.
- **indentify scientific misconduct** or even fraud by exposing statistical inconsistencies in datasets and inferences.



source: Retraction Watch

"I became increasingly cavalier and careless. I was tired of sitting at the dining table for long evenings on end, typing data into the computer. I made spreadsheet after spreadsheet to present my research ideas to the world as clearly as possible, and I made spreadsheet after spreadsheet with the data for the studies that I did with my colleagues and students. I was in a hurry, and in my haste I made mistakes. I copied rows of data from one study to another, which meant that the second dataset looked suspiciously like the first."

-Diedrick Stapel (2014) Faking Science: A True Story of Academic Fraud (trad. Nicholas J. L. Brown)

2.5.2 Peculiarities in the data matrix

For the determination of possible fraud in a publication, the original datasets were an important source of information. They enabled the Committees to compare the data with the research material and the information in the publications in order to identify discrepancies in the design and elaboration of the study (Section 2.5.1). The datasets provided could also be re-analysed in order to check the results reported in the publications (Section 2.5.3). Above all, inspection of the data matrices can reveal patterns that may at least raise suspicion of fraud. The fact that data are copied from one experiment to another is an important proof that fraud has been committed. Where there was reason for doubt, the scores or score distributions in the various datasets were scrutinized for improbable similarities, which would indicate copying. In general,

- Levelt Committee, Noort Committee, Drenth Committee (2012) Flawed science: The fraudulent research practices of socialpsychologist Diederik Stapel



> Gac Sanit. Jul-Aug 2019;33(4):356-360. doi: 10.1016/j.gaceta.2018.01.009. Epub 2018 Jun 5.

Misconduct as the main cause for retraction. A descriptive study of retracted publications and their authors

Isabel Campos-Varela ¹, Alberto Ruano-Raviña ²

Affiliations + expand

PMID: 29776690 DOI: 10.1016/j.gaceta.2018.01.009

Method: A descriptive cross-sectional study was designed including all retracted publications from January 1st, 2013-December 31st, 2016 indexed in PubMed. The causes of retraction were classified as: data management, authorship issues, plagiarism, unethical research, journal issues, review process, conflict of interest, other causes, and unknown reasons. Then, misbehaviour was classified as misconduct, suspicion of misconduct or no misconduct suspicion.

Results: 1,082 retracted publications were identified. The retraction rate for the period was 2.5 per 10,000 publications. The main cause of retraction was misconduct (65.3%), and the leading reasons were plagiarism, data management and compromise of the review process. The highest proportion of retracted publications corresponded to Iran (15.52 per 10,000), followed by Egypt and China (11.75 and 8.26 per 10,000).

Conclusions: Currently, misconduct is the main cause of retraction. Specific strategies to limit this phenomenon must be implemented. It would be useful to standardize reasons and procedures for retraction. The development of a standard retraction form to be permanently indexed in a database might be relevant.



> Gac Sanit. Jul-Aug 2019;33(4):356-360. doi: 10.1016/j.gaceta.2018.01.009. Epub 2018 Jun 5.

Misconduct as the main cause for retraction. A descriptive study of retracted publications and their authors

Isabel Campos-Varela ¹, Alberto Ruano-Raviña ²

Affiliations + expand

PMID: 29776690 DOI: 10.1016/j.gaceta.2018.01.009

Reasons for retraction are shown in Table 2. <u>Plagiarism</u> and data processing related issues were the most common reasons for retraction of publications, with 354 (32.7%) and 352 (32.5%) manuscripts retracted, respectively. Significant differences appeared when analyzing the proportion of misconduct within categories, as misconduct represented a 100% in the case of plagiarism, and only a 36.6% were certainly retracted because of misconduct in the case of data process. Review process compromised represented the third group in order of frequency, comprising a 14% of the total of retracted publications, again with a 100% accounting for misconduct. Authors' issues represented a 5.9% of the retracted papers.



> Gac Sanit. Jul-Aug 2019;33(4):356-360. doi: 10.1016/j.gaceta.2018.01.009. Epub 2018 Jun 5.

Misconduct as the main cause for retraction. A descriptive study of retracted publications and their authors

Table 2 Reasons for retraction

Isabel Campos-Varela ¹, Alberto Ruano-Raviña ²

Affiliations + expand

PMID: 29776690 DOI: 10.1016/j.gaceta.2018.01.009

Table 2. Reasons for retraction and proportion of misconduct by category.

Reason of retraction	Articles, n (%) 354 (32.7)	Misconduct, n (%)	
Plagiarism		Yes	354 (100)
		No	0
		Uncertain	0
Data	352 (32.5)	Yes	129 (36.6)
		No	1 (0.3)
		Uncertain	222 (63.1)
Review process compromised	152 (14.1)	Yes	152 (100)
		No	0
		Uncertain	0

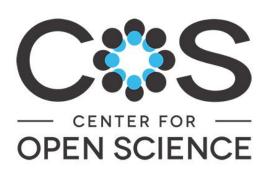


"Preregistration is the process of specifying key study and analysis details and decisions before **conducting the experiment.** The main goal of preregistering one's research is to make it easier for readers (and yourself) to distinguish between what the you set out to do (confirmation) and what was discovered along the way (exploration). Both are vital to science, but conflating these two types of work can lead to misinterpretation of the context of any claim. Preregistration prevents us from tricking ourselves and allows the argument to have meaning."

Preregistration: A Plan, Not a Prison

Alexander DeHaven

https://www.cos.io/blog/preregistration-plan-not-prison





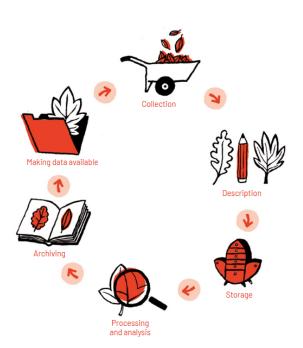


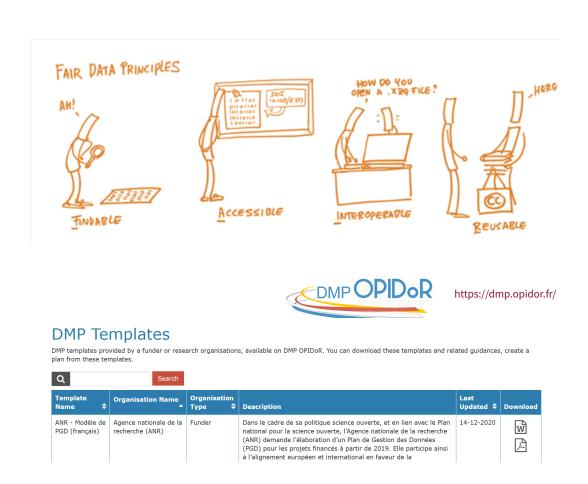
"Registered Reports is a publishing format that emphasizes the importance of the research question and the quality of methodology by conducting peer review prior to data collection. High quality protocols are then provisionally accepted for publication if the authors follow through with the registered methodology."





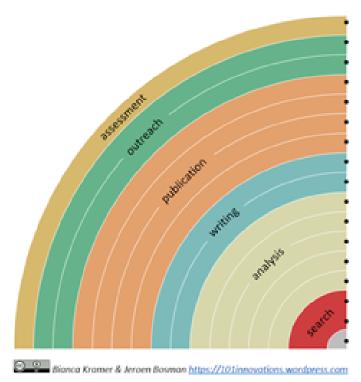
Good data management is useful for you and for others. It makes it easy to find your data and make them accessible and reusable by others. At the end of the project, it facilitates the archiving and dissemination of datasets.







You can make your workflow more open by ...



adding alternative evaluation, e.g. with altmetrics communicating through social media, e.g. Twitter sharing posters & presentations, e.g. at FigShare using open licenses, e.g. CCO or CC-BY publishing open access, 'green' or 'gold' using open peer review, e.g. at journals or PubPeer sharing preprints, e.g. at OSF, arXiv or bioRxiv using actionable formats, e.g. with Jupyter or CoCalc open XML-drafting, e.g. at Overleaf or Authorea sharing protocols & workfl., e.g. at Protocols.io sharing notebooks, e.g. at OpenNotebookScience sharing code, e.g. at GitHub with GNU/MIT license sharing data, e.g. at Dryad, Zenodo or Dataverse pre-registering, e.g. at OSF or AsPredicted commenting openly, e.g. with Hypothes.is using shared reference libraries, e.g. with Zotero sharing (grant) proposals, e.g. at RIO





2. New practices, new challenges.

- What is predatory publishing?
- A spectrum of editorial and commercial behaviors.
- A plethora of fraudulent activities.



What is predatory publishing?

"... Publication that proritize self-interest at the expense of scholarship and are characterized by false or misleading information, deviation from best editorial and publication practices, a lack of transparency, and/or the use of aggressive and indiscriminate solicitation practices."

(Grudniewicz et al., 2019)





1163 predatory publishers on

Beall's list in 2017

Last numbers before he closed his website.

10 000 predatory journals in 2018

Björk & al. 2020.

420000 predatory articles in 2014

Shen & Björk, 2015.

These are old numbers - we may be around 15 000 predatory journals in 2023.



Main characteristics of predatory publishers/journals

They provide no or only peripherical peer-review.

They ignore editorial services for publication

They often adopt deceptive tactics to entice submissions and compel payments



Historical reasons for the emergence of predatory publishing

Serials crisis in the 1980s/90s followed by the rise of APC-based gold OA in the 2000s, then followed by a widespread funding of APC-related costs by institutional actors in the 2000s/2010s.

Globalization starting in the 1980s and accelerating in the 1990s with **the Internet**, the digitalization of academic publishing, but also **western academic practices being imported in developing countries (***publish or perish***).**

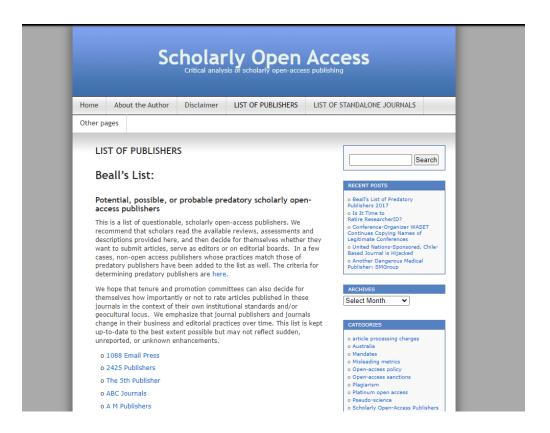




Beall's list (2010-2017)



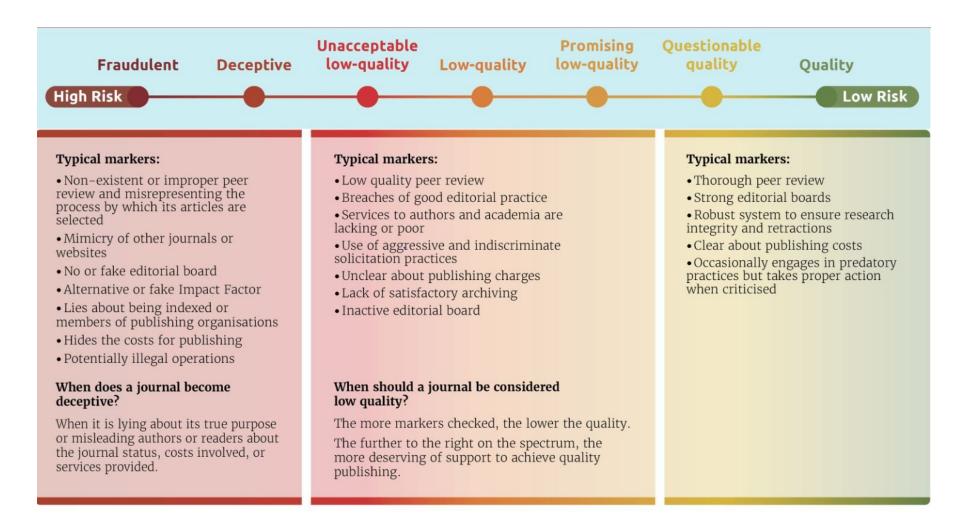
Jeffrey Beall
Librarian and associate professor at
the University of Colorado Denver
at the time he established the list.



https://scholarlyoa.com/publishers/



The complicated definition of predatory publishing: the "grey" zone





It is not just journals: predatory conferences exist, too.



https://medium.com/@westwise/ive-been-invited-to-a-super-predatoryconference-and-i-have-to-share-371ae7db6a72



https://medium.com/gradbunker/how-to-identify-predatory-conferences-688f1d4c4d82



Inside a "Fake" Conference: A Journey Into Predatory Science



https://www.technologynetworks.com/tn/articles/inside-a-fake-conference-a-journey-into-predatory-science-321619



Another type of predatory publishing: hijacked journals



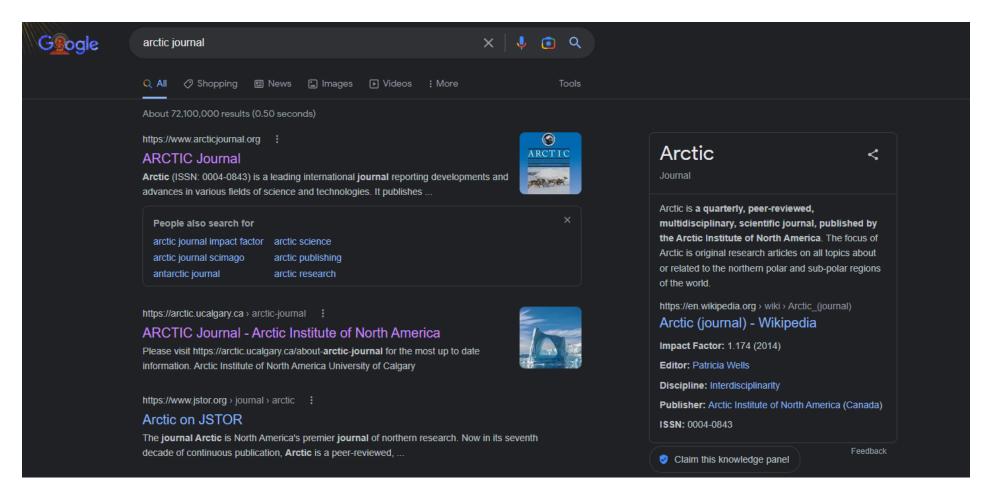
The real Arctic Journal website https://arctic.ucalgary.ca/arctic-journal



The hijacked Arctic Journal website https://www.arcticjournal.org/



Another type of predatory publishing: hijacked journals





Retraction Watch

Tracking retractions as a window into the scientific process

The Whack-a-Mole problem: Hijacked journal still being indexed in Scopus even after discovery



LINGUISTICA ANTVERPIENSIA

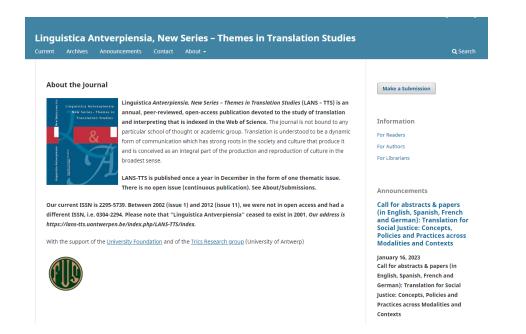
Linguistica Antverpiensia, Nouveaux thèmes de séries dans les études de traduction est une publication annuelle en libre accès, révisée par des pairs, consacrée à l'étude de la traduction et de l'interprétation qui est répertoriée dans le Web of Science. La revue n'est liée à aucune école de pensée ou groupe académique en particulier.

Sujet couvert

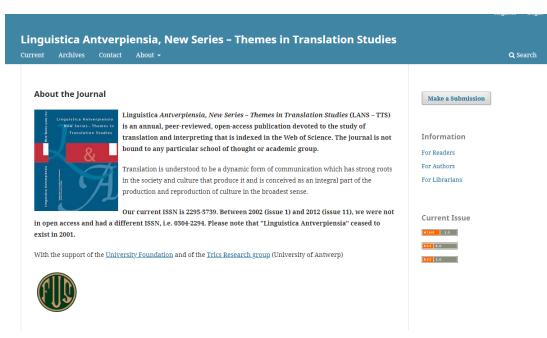
Science sociale

Sciences humaines, sociologie, éducation, science politique, droit, politique, revue sociale, arts, histoire, philosophie, anglais



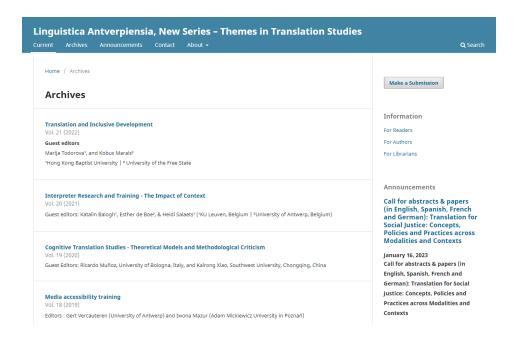


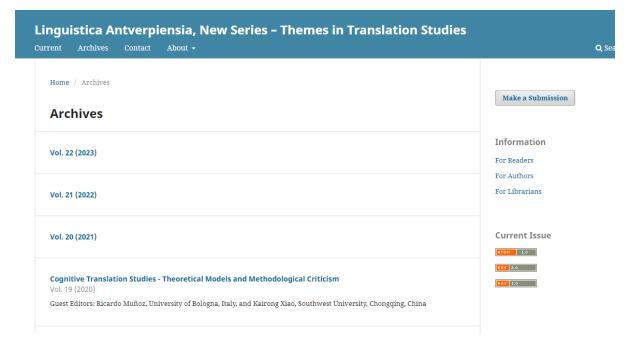
https://lans-tts.uantwerpen.be/index.php/LANS-TTS (legitimate journal website)



http://upa-editions.com/index.php/LANS-TTS (hijacked version)







https://lans-tts.uantwerpen.be/index.php/LANS-TTS (legitimate journal website)

http://upa-editions.com/index.php/LANS-TTS (hijacked version)



Home / Archives / Vol. 20 (2021)

Vol. 20 (2021)

Published: 22-12-2021

Flaubert's Correspondence as A Testimony of His Poetics

Dr Ana Markovic, MA Dragana Skrbic

120-133

DIACHRONIC PRINCIPLE OF IMPLEMENTATION OF MYTH IN LITERATURE THE REALATIONSHIP BETWEEN MYTH AND LITERATURE

Dr Ana Markovic 134-143

COMMON SPORTS INJURIES AND REHABILITATION METHODS WITH COMPLEX TRAINING FOR KABADDI PLAYERS

Dr. K. Binoy* & R. Sreejith 144-148

Lucky's Monologues and Its Chinese Translations

Li Li 149-157

The Effect of Hippotherapy and Dolphin Assisted Therapy on Neurosensorimotor Skills of Patients with Depressive Disorders

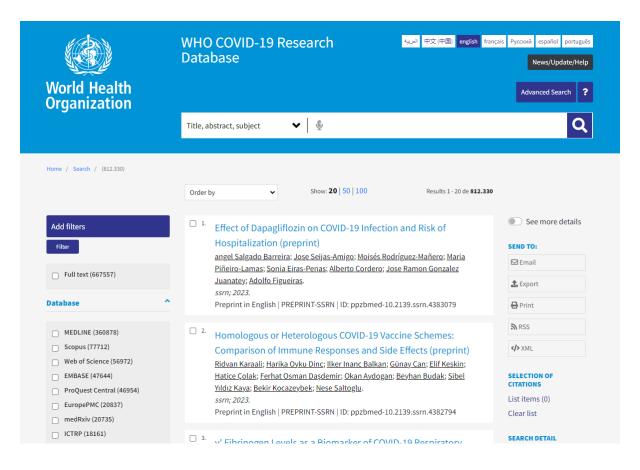
B. Kreivinienė* & D. Mockevičienė 158-174

http://upa-editions.com/index.php/LANS-TTS (hijacked version)

Focus and Scope

Each issue of *Linguistica Antverpiensia*, *New Series – Themes in Translation Studies* centres on a specific topic, in order to reflect the diversity of what is considered to be translation today, and offers an open forum for a wide range of approaches of a theoretical, methodological or descriptive nature.





Ten paper of the hijacked *Linguistica Antverpiensia ended up on the WHO Covid 19 database (they have been removed since)*







INTERNATIONAL PLATFORM FOR SEARCHING CO-AUTHORS

This platform provides an opportunity for scientists to find colleagues for joint scientific work and thus expand the geography of research, increase its value for the world scientific community. More than 11,000 scientific projects in various fields of knowledge are presented on our website.

You can take part in any study presented on the site.

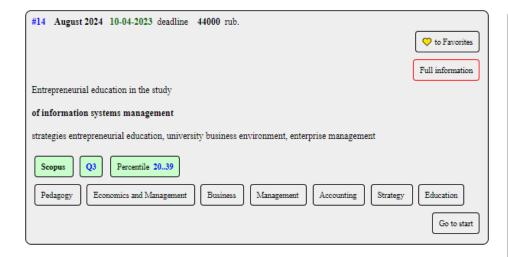
In order to avoid borrowing ideas from people not involved in the study, the details are disclosed exclusively to the participants in the study.

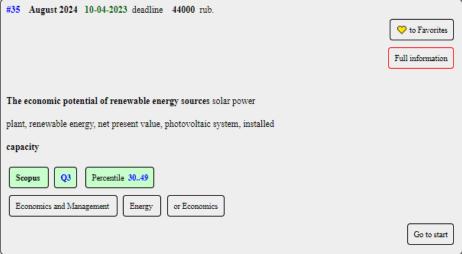
The amount required for the organization and conduct of the study is indicated separately for each work and is divided among all project participants.

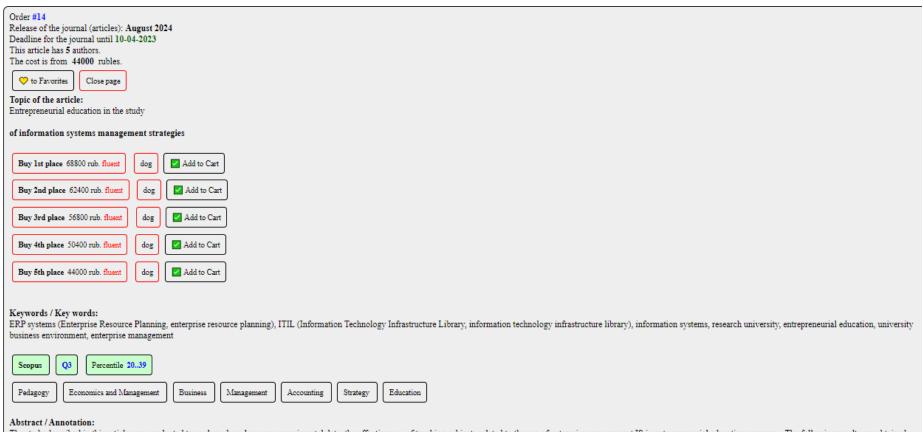
By agreeing to take part in the project, the scientist pays the specified amount of the contribution and receives a work plan, which all participants then discuss, correct and implement.

The research results are published as a scientific article in journals indexed by the Scopus or Web of Science databases.

Full information







The study described in this article was conducted to analyze, based on new experimental data, the effectiveness of teaching subjects related to the use of enterprise management IS in entrepreneurial education programs. The following results are obtained. A training module has been developed and introduced into the program of the discipline "Information Systems (IS) of Enterprise Management" to familiarize students with the concept of ITIL (Information Technology Infrastructure Library, Information Technology Infrastructure Library). This initiative was positively assessed by teachers and students.

The practical significance of the results obtained in the article lies in the fact that the proposed approach makes it possible to effectively introduce the concept of using IP to provide business services in the educational process of entrepreneurial education.



Retraction Watch

Tracking retractions as a window into the scientific

☐ July 18, 2019 👗 Adam Marcus 🕒 authorship issues

Exclusive: Russian site says it has brokered authorships for more than 10,000 researchers



A company in Russia hawks its wares

nature

NEWS FEATURE | 23 March 2021

The fight against fake-paper factories that churn out sham science

Some publishers say they are battling industrialized cheating. A *Nature* analysis examines the 'paper mill' problem – and how editors are trying to cope.

Holly Else & Richard Van Noorden

NEWS | 18 January 2023

Multimillion-dollar trade in paper authorships alarms publishers

Journals have begun retracting publications with suspicious links to sites trading in author positions.

Holly Else



Retraction Watch

Tracking retractions as a window into the scientific process

■ July 18, 2019 🔹 Adam Marcus 🕒 authorship issues

Exclusive: Russian site says it has brokered authorships for more than 10,000 researchers



A company in Russia hawks its wares

In other words, says Clarivate, the company perhaps best known for journal impact factors, the <u>Russian brokerage</u> — 123mi.ru — auctions authorship slots to researchers willing to pay hundreds of dollars for the service. Clarivate's Web of Science Group began investigating after staff at its Moscow office received a tip about the outfit, and sent a cease-and-desist letter to the operators of the service earlier this week.

According to a <u>July 17, 2019, letter</u> from the Web of Science Group to the Committee on Publication Ethics, the company found 344 articles for sale on the Russian website. Of those, 32, or 9%, are allegedly indexed on Web of Science and 303 (88%) appear in the Scopus database.



Retraction Watch

Tracking retractions as a window into the scientific process

☐ July 18, 2019 👗 Adam Marcus 🕒 authorship issues

Exclusive: Russian site says it has brokered authorships for more than 10,000 researchers



A company in Russia hawks its wares

Here's how the scheme likely works, according to Nandita Quaderi, editor in chief of Web of Science: An author, or group of authors, submits a paper and has it accepted. At that point, they submit a listing to the Russian site, offering additional authorships on the paper for a fee. The tactic exploits a vulnerability in the publishing process that allows authors to add names to manuscripts after acceptance. Such moves should raise a red flag, and have in at least a few cases.

However, at many journals, after acceptance the paper leaves the editorial department for the production offices, where workers — frequently with less experience — are unaware that it's bad practice to allow authors to add names.

"Someone working in production, right out of college, is being told by a senior professor that 'I forgot to include my co-authors," Quaderi told Retraction Watch.



Retraction Watch

Tracking retractions as a window into the scientific process

■ July 18, 2019 🛕 Adam Marcus 📦 authorship issues

Exclusive: Russian site says it has brokered authorships for more than 10,000 researchers



A company in Russia hawks its wares

Why would scientists pay to get their names onto papers they didn't write? The pressures to publish — frequently and in reputable journals — are so strong that some researchers feel the need to cut corners, cook results and take other shortcuts to get their names into the literature. In some cases, paying for the unearned privilege is simply a down payment with a potentially lucrative payoff. Institutions, and even countries, are known to offer researchers cash awards for each publication. As we <u>reported</u> in Science in 2017:

Chinese universities offer first authors more than \$43,000 for publishing a paper in Science or Nature, with the top reward for such a paper reaching a kneewobbling \$165,000.



3. The future is out there: what challenges awaits scientists?

the example of large language models and IA content generation.



The future is happening... this week, actually.

nature

NEWS 16 March 2023

GPT-4 is here: what scientists think

Researchers are excited about the AI – but many are frustrated that its underlying engineering is cloaked in secrecy.

Katharine Sanderson









The GPT-4 artificial-intelligence model is not yet widely available. Credit: Jaap Arriens/NurPhoto via Getty Images

But there is frustration in the science community over OpenAI's secrecy around how and what data the model was trained, and how it actually works. "All of these closed-source models, they are essentially dead-ends in science," says Sasha Luccioni, a research scientist specializing in climate at HuggingFace, an open-source-AI community. "They [OpenAI] can keep building upon their research, but for the community at large, it's a dead end."



Large Language Models and AI content generation in scientific research: can LLMs be authors?

NEWS | 18 January 2023

ChatGPT listed as author on research papers: many scientists disapprove

At least four articles credit the AI tool as a co-author, as publishers scramble to regulate its use.



Nurse Education in Practice

Volume 66, January 2023, 103537



Editorial

Open artificial intelligence platforms in nursing education: Tools for academic progress or abuse?

Siobhan O'Connor a 1 & M. ChatGPT b M



Large Language Models and AI content generation in scientific research: can LLMs be authors?



Science journals ban listing of ChatGPT as co-author on papers

Some publishers also banning use of bot in preparation of submissions but others see its adoption as inevitable



ChatGPT can write poems, short stories, essays and even personal advice. Photograph: Hollandse Hoogte/Rex/Shutterstock



nature

EDITORIAL 24 January 2023

Tools such as ChatGPT threaten transparent science; here are our ground rules for their use

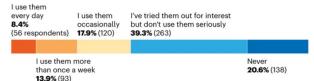
As researchers dive into the brave new world of advanced AI chatbots, publishers need to acknowledge their legitimate uses and lay down clear guidelines to avoid abuse.

Large Language Models and AI content generation in scientific research: how useful are LLMs for the scientific process?

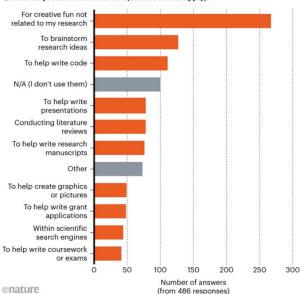
CHATBOT USERS

Most survey respondents have tried generative artificial intelligence (AI) tools, and some said they use them to help with various research-related tasks.

Q: How often do you use ChatGPT, GPT-3, DALL-E, MidJourney, Stable Diffusion, any other 'generative Al' tools, or products built using them?



Q: What do you use these tools for? (Choose all that apply).



HOW AN AI CHATBOT EDITS A MANUSCRIPT

Computational biologists Casey Greene and Milton Pividori constructed a program to get GPT-3 to edit research papers. The AI receives sections of a manuscript in turn, each with their own editing instructions (or 'prompts'). The prompts are adapted — in automated fashion — to suit the needs of each section, such as the methods or conclusion. In this example from their work, GPT-3 was fed an introduction section and the prompt below.

Prompt:

"Revise the following paragraph from the introduction of an academic paper (with the title 'An efficient not-only-linear correlation coefficient based on machine learning' and keywords 'correlation coefficient, nonlinear relationships, gene expression') so most of the citations to other academic papers are kept, the text minimizes the use of jargon, the text grammar is correct, spelling errors are fixed, and the text is in active voice and has a clear sentence structure."

Paragraph:

New technologies have vastly improved data collection, generating a deluge of information across different disciplines. This large amount of data provides new opportunities to address unanswered scientific questions, provided we have efficient tools capable of identifying multiple types of underlying patterns. Correlation analysis is an essential statistical technique for discovering relationships between variables [1]. Correlation coefficients are often used in exploratory data mining techniques, such as clustering or community detection algorithms, to compute a similarity value between a pair of objects of interest such as genes [2] or disease-relevant lifestyle factors [3]. Correlation methods are also used in supervised tasks, for example, for feature selection to improve prediction accuracy [4,5]. The Pearson correlation coefficient is ubiquitously deployed across application domains and diverse scientific areas. Thus, even minor and significant improvements in these techniques could have enormous consequences in industry and research.

GPT-3 edits:

New technologies have vasily improved data collection, generating a deluge of information across different disciplines. The increasing availability of data has opened up new possibilities for scientific exploration. This large amount of data provides new opportunities to address unanswered scientific questions, provided we have efficient tools capable of identifying multiple types of underlying patterns. Correlation analysis is an essential a useful statistical technique for discovering to uncover such relationships between variables [1]. Correlation coefficients are often used in exploratory data mining techniques, such as clustering or community detection algorithms, to compute a calculate the similarity value between a pair of two objects of interest such as, like genes [2] or disease relevant lifestyle factors related to diseases [3]. Correlation methods They are also used in supervised tasks, for example, for like feature selection, to improve boost prediction accuracy [45]. The Pearson correlation coefficient is ubiquintously deployed widely used across many application domains and diverse scientific areas disciplines. Thus Therefore, even minor and significant small improvements in these techniques could this technique can have enormous consequences in a huge impact on industry and research.

Final paragraph:

The increasing availability of data has opened up new possibilities for scientific exploration. To take advantage of this, we need efficient tools to identify multiple types of relationships between variables. Correlation analysis is a useful statistical technique to uncover such relationships [1]. Correlation coefficients are often used in data mining techniques, such as clustering or community detection, to calculate the similarity between two objects, like genes [2] or lifestyle factors related to diseases [3]. They are also used in supervised tasks, like feature selection, to boost prediction accuracy [4.5]. The Pearson correlation coefficient is widely used across many application domains and scientific disciplines. Therefore, even small improvements in this technique can have a huge impact on industry and research.

onature

Source: Adapted from ref 1

Nature 614, 214-216 (2023)

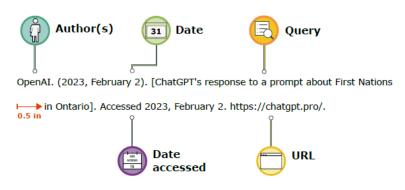
doi: https://doi.org/10.1038/d41586-023-00340-6



Large Language Models and AI content generation in scientific research: should LLMs be cited? Are LLMs sources?

APA Style References (7th Edition)

Artificial Intelligence and ChatBots



University of Waterloo



APA 7th referencing style: Generative Artificial Intelligence (AI)

A guide to using the APA 7th edition referencing system for in-text citations and reference lists.

Material Type	In-Text Citation	Reference List & Notes
Software (e.g. ChatGPT)	as generated by the generative Al program (OpenAl, 2023) According to OpenAl (2023) ChatGPT states that the relationship between language and psychology	OpenAl. (2023). ChatGPT[Computer software]. https://chat.openai.com/ If using EndNote select: Reference Type: Computer Program Programmer: OpenAl Year: 2023 Title: ChatGPT Type: [Computer software] URL: http://chat.openai.com



Thank you for your attention!

Margaux Larre-Perez, Ph.D.
Université Paris Cité Libraries
Université Paris Cité, Paris, France.

