



Large-scale flagging of Research Integrity Misconduct at Elsevier

World Conference on Research Integrity 2024

Yury Kashnitsky (Data Science)
Savvas Chamezopoulos (Data Science)
D.J. Anning (Research Integrity)





First large-scale paper mill investigation at Elsevier

We started massive paper mill investigations in fall 2022 along with the Hindawi paper mill ([RetractionWatch](#)). We flagged ~500 for manual investigation.

Example of a paper with 4 out of 5 signals:

1. ~~Presence of invalid reviewer emails — not observed~~
2. Duplicate reviewer comments – reviews by wangyantao@neepu.edu.cn are exactly or almost the same in > 200 pairs of reviewer comments
3. Too many reviews done within the same journal – wangyantao@neepu.edu.cn did 71 review for MICPRO
4. Too quick reviews done within the same journal – 47 of the reviews are done within 2 days from submission
5. Presence of tortured phrases – “image acknowledgment” (instead of “image recognition”), “recognizable proof” (seen often on Problematic Papers Screener)

Sports image detection based on FPGA hardware system and particle swarm algorithm

Jing, Hu^a  ; Xiaoqiong, Xing^b
 Save all to author list

^a Public Education Department, Anhui Finance and Trade Vocational College, Hefei, 230601, Anhui, China

^b Anhui Hefei No.168 Senior High School, Hefei, 230601, Anhui, China

12 88th percentile Citations in Scopus | 2.18 FWCI  | 17 Views count   | [View all metrics >](#)

 [View PDF](#) | [Full text options](#)  | [Export](#) 

Abstract

The action of the opposition is conveyed on a fundamental level. Moreover, it finds image detection related information on contenders and football sports Image, which can be realized in the outside image. From the force situation, one can see that image advancement is up. Therefore, this assessment development image as a thing to ponder the usage of **image acknowledgment** development. Football sports detection is dealing with the whole system. The structure configuration relies upon hardware, including a Field Programmable Gate Array (FPGA). This new computation particle swarm algorithm estimation is implemented to edge **recognizable proof**, grayscale planning, object get, target affirmation, image area development, etc., which are consolidated into the genuine need of the game video to achieve the various essentials of development **image acknowledgment**. All the while, it has set itself up as a demonstrating ground to test the suitability of the investigation framework that sees the affirmation of contenders, games affirmation, sports lead judgment, etc. Football sports Image detection results of the relevant investigations have revealed that the existence of the solution. © 2020

[Link](#)

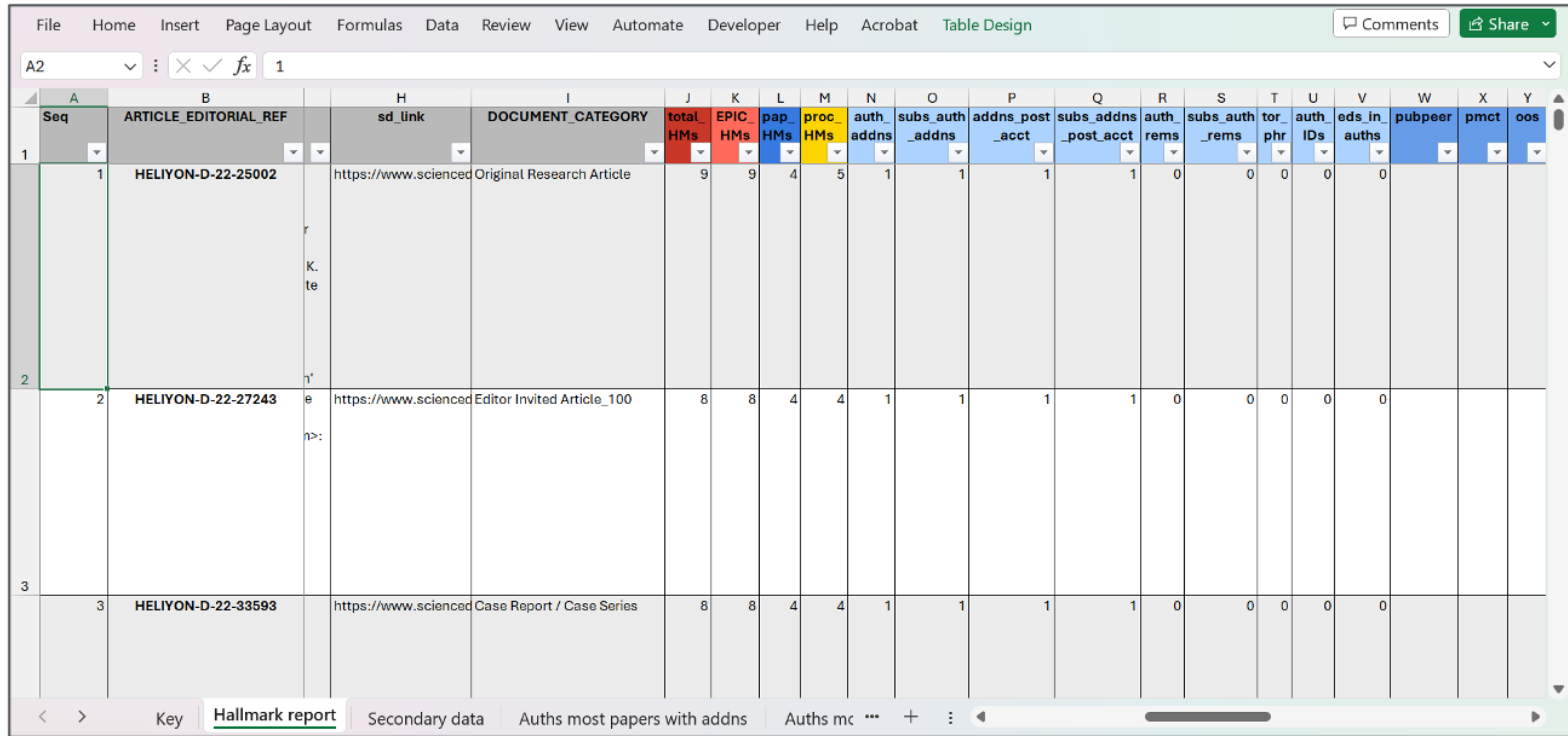
Fast-forward, in 2023, some major publishers suffered \$30-40 mln. losses due to quality concerns and Research Integrity related journal de-listings from Web of Science. [Link](#) [Link](#) [Link](#)

Research Integrity signals: main categories

Focus	Ethics Checks	Signal examples	Signals validation
Person integrity	Check scope and severity of ethics concerns about any Authors, Reviewers, or Editors associated with a paper.	<ul style="list-style-type: none"> History of retractions and withdrawals 	<p>The person-integrity signals mainly derive from established manual checks used to identify bad actors or patterns and to resolve ethics cases. They are <i>only treated as investigative leads and not used for rejections by default</i>.</p> <p>Many standard background checks are essential for determining where to focus investigative effort (or human oversight). Automating these checks is critical for enabling efficient, large-scale content screening and investigation.</p>
Authorship & manuscript integrity	Check for problems with late-stage changes to content or attribution, artificially generated text or data, suspicious collaborations, out-of-scope submissions.	<ul style="list-style-type: none"> Unapproved late additions or removals of authors 	<p>The authorship/content signals have been used to identify numerous papers associated with papermills, coercive authorship, and authorship-for-sale.</p> <p>All findings are manually checked and confirmed by investigators before corrective action is suggested to Editors.</p>
Review integrity	Check for inauthentic or inappropriate Reviewer behaviours, e.g. concerns re. speed, volume, range, etc.	<ul style="list-style-type: none"> Duplication of text across multiple reviews 	<p>The peer review signals have been used to identify journals and numerous Special Issues where the peer review process was compromised by fake or over-prolific reviewers linked to papermills and/or other coordinated misconduct.</p>
Editorial integrity	Check for potential improper handling of submissions/decisions, competing/conflicting interests, manipulating or disregarding review process.	<ul style="list-style-type: none"> Submissions accepted against, without, or with minimal review (without proper justification) 	<p>The editorial signals have been used to identify Editors and Guest Editors with major conflicts of interest, involvement in peer review manipulation, and other concerns.</p> <p>Such misconduct is high priority, since compromised Editors may handle numerous papers.</p>
Reference integrity	Check for citation manipulation by individuals and networks—including for large-scale patterns that can affect journal indexing & reputation.	<ul style="list-style-type: none"> Inappropriate or excessive citation-prompting 	<p>The citations signals have been developed and trialled across more than a dozen journals (~500,000 total papers).</p> <p>Testing on the first two cases led to rapid detection of 8 citation networks. This included mapping and evidencing a complex citation ring centred on two Associate Editors and their</p>

Investigation tools: Editorial Process Integrity Checker (EPIC)

The EPIC tool reports ~20 RI signals for up to ~ 1 mln. papers at a time. Unlike other tools, it also collects **secondary data** supporting the signals



Seq	ARTICLE_EDITORIAL_REF	sd_link	DOCUMENT_CATEGORY	total_HMs	EPIC_HMs	pap_HMs	proc_HMs	auth_addns	subs_auth_addns	addns_post_acct	subs_addns_post_acct	auth_rems	subs_auth_rems	tor_phr	auth_IDs	eds_in_auths	pubpeer	pmct	oos
1	HELIYON-D-22-25002	https://www.sciencedirect.com/doi/10.1016/j.heliy.2022.25002	Original Research Article	9	9	4	5	1	1	1	1	0	0	0	0	0			
2	HELIYON-D-22-27243	https://www.sciencedirect.com/doi/10.1016/j.heliy.2022.27243	Editor Invited Article_100	8	8	4	4	1	1	1	1	0	0	0	0	0			
3	HELIYON-D-22-33593	https://www.sciencedirect.com/doi/10.1016/j.heliy.2022.33593	Case Report / Case Series	8	8	4	4	1	1	1	1	0	0	0	0	0			

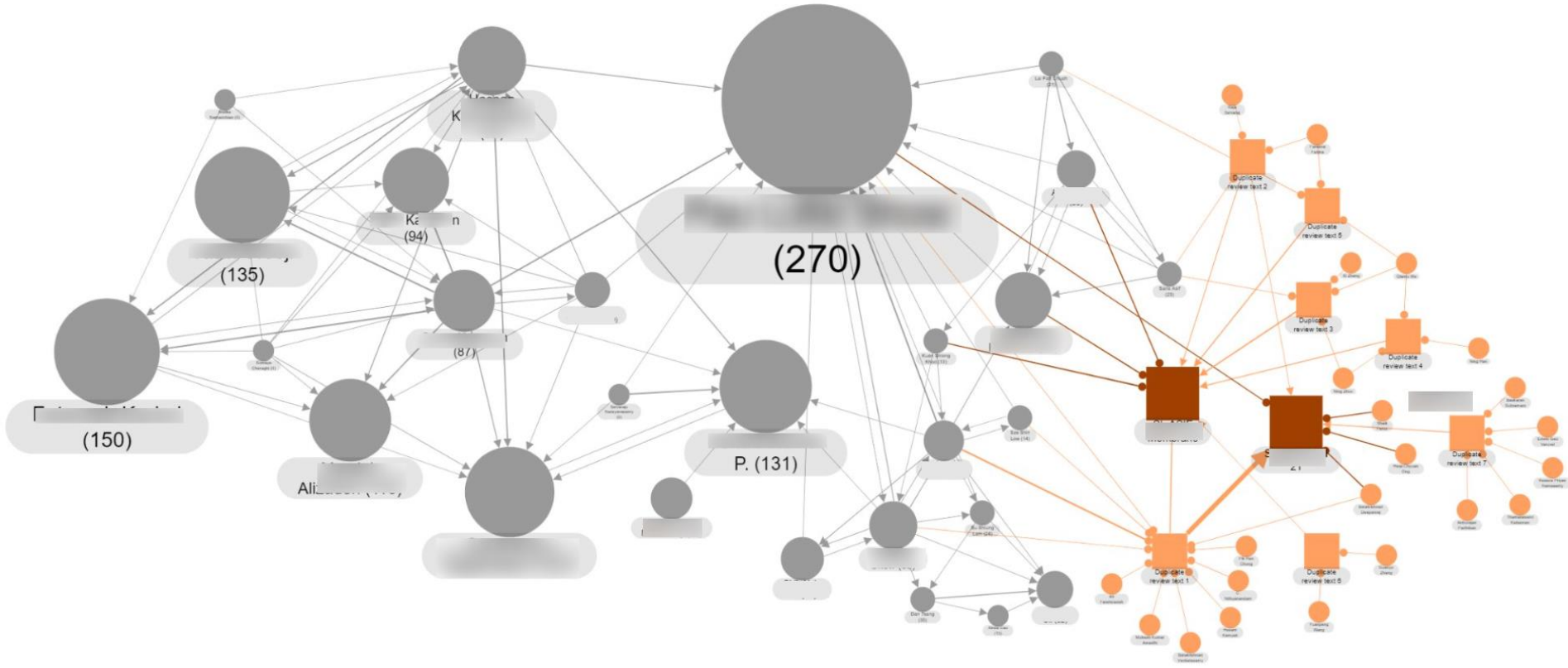
EPIC reports: secondary data

Secondary data is the *actionable evidence* associated with an integrity signal, which can quickly help an investigator or screener determine the validity and severity of a concern and select the appropriate next steps. Each batch of secondary data in EPIC was developed, tested, and tuned during live casework.

#	Example Elsevier-developed signals	Example secondary data provided to support decision-making
1	Late-stage authorship additions/removals	<ul style="list-style-type: none"> Account details of the authors added/removed. Indication of the scale of the authorship changes. Stage at which the changes were made (e.g. during revisions or post-acceptance).
2	Accepted against reviewer recommendations	<ul style="list-style-type: none"> List of reviewers and the recommendations provided with their review reports. Name and email of the handling editor.
3	Duplicate review comments	<ul style="list-style-type: none"> Full text of the duplicate review fragment(s). Number of review reports in which the fragment recurred within the dataset. Account details of the EM user(s) associated with the repetition.
4	Prolific reviewer	<ul style="list-style-type: none"> Account details of the reviewer who exceeded the plausible review output threshold. Number of first-round review reports completed within the threshold period.
5	Citation-prompting	<ul style="list-style-type: none"> Account details of the flagged reviewer. Numbers of potential prompts included in their review report. Full text of the review fragment containing the suspected prompts.

Investigation tools: Citation Analysis

The citation analysis tool helps us detect citation manipulation signals



Investigation tools: Research Integrity Outlier Tracker (RIOT)



RIOT is similar to EPIC with a difference that it detects outliers among a wide range of journals. Such a watchdog would then send alerts to RI investigators.

Currently implemented signals:

- **Author-Based:**
 - Affiliation count
 - Citation count (to Author by others)
 - Citation count (from Author to others)
 - Citation per paper count
- **Paper-based:**
 - Papers with too many citations
 - Papers with too narrow references

Investigative tools: case impacts

Integrity signals become usable and can exert an impact during cases when combined with actionable data and evidence to support further decision-making. In prototype format, the EPIC tool and the Citations Dashboards have provided insights that are helping to resolve ethics concerns in >90 large-scale, complex cases.

Case	Example issues detected using signals & secondary data	Impacts of identified issues to date
Case #1	<ul style="list-style-type: none"> • Editors handling/accepting own papers (<i>Editorial COIs</i>) • Unapproved late-stage changes to papers & authorship (<i>Author additions/removals</i>) • Reviewer recommendations & concerns disregarded (<i>Accepted without/with minimal/against review</i>) 	<ul style="list-style-type: none"> • Cancellation of Society journal contract. • Removals of editors involved in misconduct. • Possible reputational damage for titles with systematic issues.
Case #2	<ul style="list-style-type: none"> • Recurring editorial COIs and handling of own papers (<i>Editorial COIs</i>) • Systematic citation-stacking (<i>Citation-prompting</i>) • Suspected authorship-for-sale (patterns of <i>Author additions</i>) • Coordinated manipulation of peer review process (<i>Generic/superficial and duplicate reviews</i>) 	<ul style="list-style-type: none"> • Removal of longstanding EiC, Associate Editors, and Editorial Board Members. • Full historical review of the editorial records of those involved in misconduct, with appropriate corrective actions in process.
Case #3	<ul style="list-style-type: none"> • Citation-rings enabled by several Guest Editors (<i>Citation outliers</i>) • Coordinated network of review manipulation (e.g., <i>Duplicate reviews</i>) • High proportion of out-of-scope content in SIs (<i>Scope-match assessment</i>) 	<ul style="list-style-type: none"> • Expected retractions or other corrections of ~1.8% of journal content due to COIs, authorship issues, review problems. • Cancellation of multiple Special Issues, mainly for coordinated misconduct by Guest Editors and Reviewers.
Case #4	<ul style="list-style-type: none"> • Audit of batch of Special Issues both before and after publication, due to papermill-style reviews (<i>Generic/superficial and duplicate reviews</i>) • Undeclared use of artificially generated text (e.g., <i>Tortured phrases</i>) • Coordinated citation manipulation by prolific Guest Editors (<i>Citation-prompting, Coercive citation</i>) 	<ul style="list-style-type: none"> • Several Special Issues undergoing retractions/withdrawals. • Enhanced vetting protocols developed for SIs, including detailed cross-checking of editorial patterns and behaviours across EPIC reports.



ELSEVIER

Contacts

dj.anning@elsevier.com

s.chamezopoulos@elsevier.com

