

Review

Not peer-reviewed version

Analysis of Gross Methodological Errors in Webometrics Ranking Data (July 2025) Using Yemeni Universities as a Case Study

[Saddam Al-Sifi](#) *

Posted Date: 2 September 2025

doi: 10.20944/preprints202509.0019.v1

Keywords: webometrics ranking; research organization registry (ROR); data integrity; methodological errors; university rankings; Yemeni universities



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Review

Analysis of Gross Methodological Errors in Webometrics Ranking Data (July 2025) Using Yemeni Universities as a Case Study

Saddam Al-Slfi

Director of Development and Computing Computer and Information Technology Department Queen Arwa University, Sana'a, Yemen; saddamalslfi@qau.edu.ye

Abstract

This paper presents a critical analysis of the dataset for the Webometrics Ranking of World Universities, July 2025 edition, as published by Aguillo (2025) [1]. Focusing on Yemeni universities as a case study, the analysis reveals multiple patterns of gross errors in the assignment of Research Organization Registry (ROR) identifiers. The study documents "chained errors," where identifiers are incorrectly swapped among several universities, in addition to cases of complete omission of universities or the failure to assign their correct, existing identifiers. All findings presented herein are based on the published data from the specified source. These profound methodological flaws raise fundamental doubts about the data validation mechanisms of the Webometrics ranking and directly impact the fairness and credibility of its results.

Keywords: webometrics ranking; research organization registry (ROR); data integrity; methodological errors; university rankings; Yemeni universities

1. Introduction: The Impact of Global Rankings and the Importance of Data Integrity

Global university rankings have become profoundly influential tools in the higher education sector, affecting policy decisions, funding allocations, prospective student choices, and the academic reputation of institutions [2]. Among these, the Webometrics Ranking of World Universities, produced by the Cybermetrics Lab, stands out for its extensive scope, covering nearly 32,000 higher education institutions worldwide. The ranking aims to promote open access to the knowledge generated by universities and provides unique coverage of many institutions in the "Global South" often overlooked by other rankings [3].

However, web-based rankings, including Webometrics, have not been immune to academic criticism. The literature suggests that such rankings may contain methodological biases, tending to favor older, well-resourced institutions, while universities in developing nations are negatively affected by factors like the digital divide and limited internet access. Concerns have also been raised about their reliance on quantitative metrics that may overlook qualitative aspects of academic excellence, as well as a lack of transparency in data collection and verification methodologies [4].

In an effort to address challenges of data ambiguity and accuracy, the academic community has increasingly moved towards adopting Persistent Identifiers (PIDs) as a standard solution. A prominent example is the Research Organization Registry (ROR), an open, community-led infrastructure designed to provide a unique and persistent identifier for every research organization in the world. The ROR ID disambiguates institutions with similar names and tracks them through structural or name changes, ensuring clean and consistent metadata. The adoption of ROR IDs by essential scholarly infrastructure systems like Crossref and DataCite makes their correct usage a hallmark of sound data management. Webometrics' decision to incorporate ROR IDs is not merely a technical update but an implicit declaration of commitment to accuracy and modern standards of data integrity, aimed at bolstering its credibility against existing methodological critiques.

This study’s central thesis is that although the Webometrics ranking has adopted ROR IDs to enhance accuracy, a detailed analysis of its published data for the July 2025 edition reveals a widespread methodological failure in their application. Using Yemeni universities as a ‘critical case’ to probe for systemic weaknesses, this paper presents empirical evidence of gross methodological errors. This shifts the critique from the ranking’s metrics to its fundamental data governance, demonstrating that this failure is not a mere technical glitch but a failure to meet the very standards the ranking claims to uphold.

2. Analytical Framework and Methodology

This study was designed to be replicable and transparent, relying on public data sources and a rigorous verification process. The methodological framework aims to document errors irrefutably and present them within a clear classification.

2.1. Data Sources

The analysis was based on two primary data sources:

- **Primary Data Source:** The dataset under review is the "Ranking Web of Universities (webometrics.info). July 2025 edition (Version 1)," published by Isidro F. Aguillo via the "figshare" platform. This dataset is precisely identified by its Digital Object Identifier (DOI): <https://doi.org/10.6084/m9.figshare.29588921.v1> [1].

A notable structural limitation of the published dataset is the absence of a dedicated ‘Country’ column for each listed institution. This omission complicates data sorting and verification, and can introduce ambiguity, especially when dealing with universities that have similar names across different regions.

- **Reference Verification Source:** The official Research Organization Registry (ROR) was used as the primary reference for verifying the correctness of identifiers. To ensure comprehensiveness and accuracy, verification was conducted by analyzing a full ROR database dump, publicly available via Zenodo [5]. Additional cross-verification was performed using other persistent identifier databases such as Wikidata to ensure consistency.

2.2. Scope of the Study

The scope of the analysis was intentionally limited to the sample of Yemeni universities listed in the Webometrics data file. This approach provides a focused and in-depth case study capable of revealing systematic patterns of errors in a manageable and fully documentable manner.

2.3. Classification of Errors

To systematically analyze the nature of the errors, a classification of four main types was developed based on initial observations of the data. Each type is defined as follows:

1. **Type I: Chained Errors:** The most complex pattern, where the ROR ID of university (B) is assigned to Yemeni university (A), while the correct ID for university (A) is assigned to a third university (C).
2. **Type II: Misidentification:** The direct assignment of an incorrect ROR ID belonging to another institution, often one with a similar name, to a Yemeni university.
3. **Type III: Omission of Existing IDs:** A university that officially possesses a ROR ID is listed, but its identifier field is left blank in the dataset.
4. **Type IV: Complete Omission:** The total absence of a university from the dataset despite its existence and possession of an official ROR ID.

3. Results: A Typology of Data Integrity Failures in the Yemeni University Sample

A meticulous examination of the Webometrics ranking data for the July 2025 edition reveals serious and systemic methodological flaws in the handling of ROR IDs within the Yemeni university sample. These errors have been categorized into four distinct types, illustrating the systemic nature of the problem rather than isolated, random mistakes.

3.1. Type I: Chained Errors

This error type represents the most egregious failure in data processing, creating a complex web of incorrect assignments that links unrelated institutions across continents. Table 1 illustrates this chain of errors, documented directly from the published data file.

Table 1. Documented Chained Errors in ROR ID Assignment for Yemeni Universities in the Webometrics Dataset (July 2025)

Yemeni University (A)	Incorrect ID Assigned to A (from B)	Original Owner of the ID (University B) & Country	Correct ID of A (Verified)	University Incorrectly Assigned A's ID (University C) & Country
Azal University for Human Development	https://ror.org/030chaq85	Yamaguchi College of Arts (Japan)	https://ror.org/02zv8ns48	Escuela Superior de Guerra Naval (Peru)
University of Saba Region	https://ror.org/01p0vsd73	CETT Barcelona School... (Spain)	https://ror.org/01nd4jr17	Biwako Gakuin University (Japan)
Al-Qalam University for Humanities and Applied Sciences	https://ror.org/02v4dqa55	University of Chakwal (Pakistan)	https://ror.org/02v6vgb19	Universitas Aufa Royhan (Indonesia)
Hajjah University	https://ror.org/01s7pfd33	Alfraganus University (Uzbekistan)	https://ror.org/01rpcwa78	Universitas Al-Irsyad Cilacap (Indonesia)
Al-Andalus University for Science & Technology	https://ror.org/04jy6j173	Politeknik Pariwisata NHL... (Indonesia)	https://ror.org/04mnyr134	Sonoda Women's University (Japan)

Source: Analysis of data published by Aguillo (2025) [1] and cross-referenced with the ROR registry [5].

3.2. Type II: Misidentification

This error occurs when two different institutions are confused due to name similarity—precisely the type of problem ROR IDs are designed to solve.

- **Al Jazeera University, Ibb:** According to the data file [1], this university was assigned the ID <https://ror.org/00basmr24>. Upon verification, this ID officially belongs to Aljazeera Private University in Syria, a clear case of institutional misidentification.

3.3. Type III: Omission of Existing IDs

In these cases, the a shows that some Yemeni universities with official, registered ROR IDs were included in the ranking without their identifiers being listed, indicating a failure to perform a basic check against the ROR database.

- **Al-Nasser University:** Its official ID <https://ror.org/02rsbbb97> was omitted.
- **Emirates International University:** Its official ID <https://ror.org/03j6pc929> was omitted [1].

3.4. Type IV: Complete Omission from the Ranking

This type of error represents a failure at the initial stage of compiling the list of institutions, where universities were entirely excluded from the dataset despite their existence and possession of an official ROR ID.

- **Iman University:** This university, which holds the official ROR ID <https://ror.org/04ajy5s58>, is completely absent from the Webometrics dataset published by Aguillo (2025) [1].

Table A1 in the Appendix presents the dataset with manually corrected ROR identifiers to accurately match each institution. It is crucial to note that the ranking scores shown remain the original data published by Webometrics. A definitive correction of the ranks would require a re-calculation by the Cybermetrics Lab based on these accurate institutional identities.

4. Discussion: Implications for Ranking Credibility and Fairness

The methodological failures documented in the results are compounded by a recent and alarming lack of transparency regarding the ranking's online presence. The official Webometrics portal (<https://webometrics.info>), which historically published detailed ranking data, has ceased its updates since the July 2025 edition, with no notice of maintenance, migration, or discontinuation. Compounding this ambiguity, a new and seemingly unaffiliated website has emerged at <https://webometrics.org/>. This new portal uses the same branding but carries a disclaimer explicitly stating it is an "independent university ranking platform and is not affiliated with the former Webometrics[.info] website or the Consejo Superior de Investigaciones Científicas (CSIC)." This schism creates profound confusion and further erodes the trustworthiness of the Webometrics brand, leaving users and institutions unable to determine the official source of data or the status of the ranking itself.

The patterns of errors documented in this analysis transcend mere individual oversights to signify deep methodological failures in the data collection and validation processes at Webometrics. These findings, derived directly from the ranking's official data, have serious implications for the credibility and fairness of the ranking, both locally and globally.

The existence of "chained errors," in particular, strongly suggests something beyond human data entry mistakes. The interconnected nature of these errors, where IDs are swapped between institutions in Yemen, Japan, Peru, and Spain, implies a catastrophic failure in an automated process, such as an inaccurate name-matching algorithm or the merging of spreadsheets without using key fields for verification. This pattern of error reveals an absence of basic quality controls and sound data governance mechanisms at the Cybermetrics Lab; the severity and systemic nature of these flaws make it difficult to dismiss them as mere "teething problems" of a new system.

It is deeply ironic that a system like ROR, designed specifically to eliminate ambiguity and ensure accuracy, has been used in a way that creates new layers of chaos and distortion. This superficial adoption of persistent identifier technology, without a commitment to the underlying principles of data integrity it represents, not only devalues the technology but also misleads the academic community, who might assume that the use of ROR is a guarantee of accuracy. The problem is not a scarcity of correct data, but a clear failure to use it. ROR IDs are universally accessible via APIs or full database downloads. The fact that this readily available reference data was not used to validate the assignments indicates either a significant lack of technical competence or a gross neglect of accuracy standards.

The damage caused by these errors is not theoretical but tangible, affecting the institutions involved. When a Yemeni university is incorrectly linked to an institution in Pakistan, all associated bibliometric and cybermetric indicators are skewed, artificially inflating or deflating its perceived research performance. This digital misrepresentation harms the universities' reputations, their ability to attract talented students and researchers, and their standing in the global academic community.

Although this study focuses on a specific sample from Yemen, the systemic nature of the identified flaws makes it highly probable that similar errors exist in the data for universities in other countries, especially those in the Global South that may lack the resources to audit and challenge their data. These findings, therefore, do not just question the accuracy of the ranking for Yemeni universities but cast a shadow of doubt over the integrity and reliability of the entire global Webometrics ranking. If such fundamental errors exist in a simple process like ID matching, how can the more complex calculations that the ranking relies on be trusted?

5. Conclusion and Recommendations

This critical review demonstrates, with direct and documented evidence from the analysis of the Yemeni university sample in the published Webometrics data for the July 2025 edition, that there are serious and varied methodological flaws in the handling of ROR IDs. These failures, ranging from misidentification to complex chained errors, fundamentally undermine the ranking's credibility and raise serious questions about its data validation mechanisms.

Based on these findings, the following recommendations are made:

Recommendations for the Webometrics Ranking Administration (Cybermetrics Lab):

1. **Immediate Retraction and Correction:** The flawed July 2025 dataset should be immediately retracted, and a corrected version should be published after a thorough verification of all ROR IDs.
2. **Comprehensive Audit:** A full and transparent audit of all global data in the ranking should be conducted to identify similar errors in ROR ID assignments for universities in other countries.
3. **Methodological Transparency:** A detailed report should be published explaining precisely how data is collected, validated, and how persistent identifiers are integrated, outlining the quality assurance procedures in place to prevent such errors from recurring.
4. **Process Re-engineering:** Robust automated validation protocols should be adopted. This includes using the ROR API to programmatically verify each ROR ID and implementing fuzzy name matching algorithms with country-level filtering to flag potential mismatches before publication.

Recommendations for Affected Academic Institutions:

1. **Official Challenge:** Yemeni universities and other institutions that suspect inaccuracies in their data should use the evidence documented in this study to formally contact the Webometrics administration and demand immediate correction of their data.
2. **Proactive ID Management:** Universities are encouraged to claim and regularly update their official profiles in the ROR registry to ensure their metadata is accurate and authoritative at the source.

Recommendations for the Academic Community:

1. **Critical Scrutiny:** Researchers, policymakers, and funding bodies are called upon to exercise a higher degree of critical scrutiny of all university rankings and to demand higher standards of transparency and data integrity from ranking providers.
2. **Promote PID Adoption:** Emphasize the importance of consistent use and advocate for the adoption of open persistent identifiers like ROR across all academic systems to build a more reliable and interconnected research ecosystem.

Appendix A. Corrected Webometrics Ranking for Yemeni Universities (July 2025)

The following table presents the ranking of Yemeni universities based on the Webometrics data, with corrected ROR IDs assigned where applicable.

Table A1. Corrected Webometrics Ranking for Yemeni Universities (July 2025).

University Name	Corrected ROR ID	Country Rank	World Rank
Sana'a University	https://ror.org/04hcva32	1	3894
University of Science and Technology	https://ror.org/05bj7sh33	2	4703
Ibb University	https://ror.org/00fhcx56	3	5171
Thamar University	https://ror.org/04tsbkh63	4	5226
Taiz University	https://ror.org/03jwcx96	5	5279
University of Aden	https://ror.org/02w043707	6	5373
Hadhramout University	https://ror.org/02kv0px94	7	5747
Hodeidah University	https://ror.org/05fkpm735	8	5890
Al-Razi University	https://ror.org/04rrnb020	9	6983
Queen Arwa University	https://ror.org/03ygqq617	10	7812
Amran University	https://ror.org/055y2t972	11	8215
Saba University	https://ror.org/051kvx87	12	9208
Albaydha University	https://ror.org/0505vtn61	13	9551
Aljanad University for Science and Technology	https://ror.org/05ngpb650	14	10376
Sana'a Community College		15	11229
Seiyun University		16	11960
Azal University for Human Development	https://ror.org/02zv8ns48	17	12665

Table A1. Cont.

University Name	Corrected ROR ID	Country Rank	World Rank
Lebanese International University	https://ror.org/027anng05	18	14688
Al-Ahgaﬀ University	https://ror.org/040jyv820	19	16232
Modern Specialized University	https://ror.org/01n0j2c74	20	16466
University of Science and Technology, Sanaa	https://ror.org/0520msa48	21	18694
University of Holy Qur'an and Islamic Sciences Hadramaut		22	20756
Al-Nasser University	https://ror.org/02rsbbb97	23	21687
University of Saba Region	https://ror.org/01nd4jr17	24	21762
Future University Yemen		25	21823
Al-Saeeda University	https://ror.org/05v0zt272	26	21958
University of Modern Sciences	https://ror.org/01crf4k59	27	22408
Yemen University		28	22777
Yemenia University	https://ror.org/022jg8f66	29	22927
Yemen College of Middle Eastern Studies		30	23120
National University	https://ror.org/046s04e65	31	23391
Al Jazeera University Ibb		32	24494
Al-Saeed University	https://ror.org/04gkkrw50	33	24718
21 September University of Medical and Applied Sciences	https://ror.org/05b8hjk91	34	25736
Arabian University Sana'a		35	26185
Al-Rayan University	https://ror.org/01ktn5v16	36	27123
Yemen Academy for Graduate Studies	https://ror.org/04vt2s547	37	27266
Emirates International University	https://ror.org/03j6pc929	38	27270
Al Hikma University	https://ror.org/02gljdz81	39	27569
Yemeni Jordanian University	https://ror.org/02ggf1973	40	27859
Ibn Khaldoun University		41	28212
International University of Technology Twintech	https://ror.org/03xztvt08	42	28384
Al-Qalam University for Humanities and Applied Sciences	https://ror.org/02v4dqa55	43	29184
Mahrah University	https://ror.org/05c1b7t53	44	29384
Civilization University	https://ror.org/035ky6r06	45	29818
Dar Al Salam International University for Science and Technology Sana'a		46	29822
Modern Specialized College for Medical and Technical Sciences		47	29822
Jiblah University for Medical and Health Sciences		48	30065
Knowledge & Modern Sciences University		49	30065
Hajjah University	https://ror.org/01rpcwa78	50	30306
Sa'ada University	https://ror.org/03xv17r49	51	31294
Al-Ataa University for Science and Technology		52	31294
Alandalus University For Science & Technology	https://ror.org/04jy6j173	53	31534
Genius University for Sciences & Technology	https://ror.org/03qx2bq13	54	31534
Aljeel Aljadeed University	https://ror.org/04kjeyv82	55	31684

Source: Data compiled from Aguillo (2025) [1] and corrected using the official ROR registry [5]. Universities without a ROR ID listed do not currently have one assigned in the public registry.

References

1. Isidro F. Aguillo. Ranking web of universities (webometrics.info). july 2025 edition (version 1). *figshare*, 2025. doi: 10.6084/m9.figshare.29588921.v1.

2. M. A. Fauzi, C. N. L. Tan, and M. H. Ngerng. University rankings: A review of methodological flaws. *Issues in Educational Research*, 30(1):79–96, 2020.

3. F. Kinyanjui. Webometrics ranking of Universities: fallacy or reality. *African Journal of Science, Technology and Social Sciences*, 3(1):1–8, 2024.
4. United Nations University. Rethinking Quality: UNU-convened Experts Challenge the Harmful Influence of Global University Rankings. November 2023. Retrieved from <https://unu.edu/press-release/rethinking-quality-unu-convened-experts-challenge-harmful-influence-global-university>.
5. ROR Community. Research Organization Registry (ROR) Data Dump. Zenodo, 2025. doi: 10.5281/zenodo.15925254.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.