

**Toluwase Victor Asubiaro**  
**Western University, London, ON, Canada**  
**University of Ibadan, Ibadan, Nigeria**

# Digital archiving by Nigerian and foreign authors in a low resource context: A Content Analysis of publications on Natural Language Processing of Nigerian Languages

## **Abstract**

This study investigated if there is a difference in the number of articles, datasets and computer codes that foreign and Nigerian authors of scientific publications on natural language processing (NLP) of Nigerian languages deposited in digital archives. Relevant articles were systematically retrieved from Google, Web of Science and Scopus. Authorship type and data archiving information was extracted from the full text of the relevant publications. Result shows that papers with foreign authorship (80.4%) published their articles in non-commercial repositories, more than papers with Nigerian authorship (55.3%). Similarly, few papers with foreign authorship deposited research data (19.1%) and computer codes (10.4%), while none of the papers with Nigerian authorship did. It was recommended that librarians in Nigeria should create awareness on the benefits of digital archiving and open science.

**Keywords:** data sharing, open science, digital archiving, natural language processing

## **1. Introduction**

Considering challenges such as poor funding, poor access to world class scientific publication databases and low visibility of research (Asubiaro, 2017, 2019; Kari & Baro, 2016, Bakare, Munir, & Bello-Mojed, 2014, Pouris & Ho, 2014), facing research in Nigeria and many other African countries, the open scholarship model of scientific communication is potentially a silver linen, if it is judiciously harnessed. The potential benefits of the open scholarship model of research communication include increased access to knowledge outputs from other parts of the world through the open access (OA) model, better educational resources through the open education resources (OER), more access to digital resources through the open source software (OSS) platforms. Other types of the open scholarship model including openLabs, open workshops and open science have the potentials to provide access to more and better resources for research. Research visibility for research from Africa through compliance and implementation of open repositories in Africa and improved research quality through open data (OD) and OA (McKiernan, 2017; Pinto, 2019; Raju et al., 2015; Tennant et al., 2019) is another possibility. It is also evident that open scholarship is relevant to alleviating the resource-scarcity in data intensive fields in Africa, such as natural language processing (NLP) of its languages which have little data and software to work with.

Digital archiving, which is the practice of sharing digital copies of research outputs (publications, research data, etc) to online archives or repositories so that they could be accessed without cost, is

pivotal to the open scholarship practice. With all the potential benefits of the open scholarship to Africa's science, there is a research gap on how it has been harnessed by researchers in the continent. This study, therefore, sets out to investigate how Nigerian authors, in comparison to their foreign counterparts, archive publications, research data and computer codes of native Nigerian languages' NLP studies in non-commercial repositories. Scarcity of computing resources for the Nigerian languages is a major challenge for the NLP of the languages.

## 2. Methodology

Bibliographic data of publications on the NLP of the top ten languages from Nigeria were retrieved from Google search engine, Web of Science and Scopus databases using the search strategies in Asubiaro and Igwe (2020) and adopted in Asubiaro (2020). After the removal of non-relevant articles, from the retrieved articles from the three online sources, 99 articles were found useful. In addition, references of the 99 articles were searched for relevant articles, which returned 15 relevant articles. All the 114 articles were considered for analysis; one article was excluded because it was newly published and its availability on repositories could not be evaluated. The internet was searched using Google and Google Scholar search engines for the availability of publications on non-commercial repositories.

Authors' country of affiliation could be Nigeria, outside Nigeria or mixed. Where all authors are affiliated to Nigerian institutions, authorship type was labeled "Nigeria". Where all authors are affiliated to institutions outside Nigeria, authorship type was recorded as "foreign". Where one or more authors are from Nigeria while others are affiliated to institutions outside Nigeria, authorship type was recorded as "mixed". An article was considered open on the non-commercial repository if the full text, which could be the pre-print, authors' accepted manuscript, or the published article, is accessible on at least a non-commercial repository. Computer code or research data was also considered open if it is accessible and available on a non-commercial repository.

## 3. Results

The number of articles in which all the authors are all foreigners (44.74%) surpassed the number of articles in which all the authors are from Nigeria (41.22%) and mixed authorship (14.03%). Most (56.14%) of the articles were originally published in open access journals. In addition, most of the articles (65.79%) were made available in at least a repository regardless of their original status as open or closed access articles. The results are presented under two main themes: open access and open data. Some articles were excluded in the open data analysis because data and computer codes use were not relevant to the studies.

### 3.1 Open Access: Publications' Availability on Non-commercial Repositories

All the 114 articles were included in this part of the study. Articles on computer processing of Nigerian languages appeared more in foreign information repositories (IRs) (31.58%) than in local IRs (2.63%). Table 1 shows that 55.3% of the articles that were published by Nigerian authors were made available on at least one non-commercial repository as against 80.4% of the articles that were published by foreign authors.

*Table 1: Digitally archived Publications by Nigerian and non-Nigerian authors*

	Authorship type			Total
	Nigerian	Mixed	Foreign	

Academia.edu	9 (19.1%)	5 (31.32%)	5 (9.8%)	19(16.5%)
Researchgate.net	17 (36.2%)	7 (43.75%)	30(58.8%)	55 (47.8%)
Others	15 (31.9%)	4 (25%)	28 (54.9%)	47(40.9%)
At least one	26 (55.3%)	8 (50%)	41 (80.4%)	70(66.1%)

### 3.2 Open Data: Published Research Data and Computer Codes

Of all the 114 articles, only 104 articles were found relevant and included in the section of the study. Table 2 shows that research data (corpora etc) was published in none of the papers that were authored by scientists from Nigeria, while research data was published in 19.1% of foreign authored papers, than the mixed authored papers (11.8%).

*Table 2: Digitally archived research data by Nigerian and non-Nigerian authors*

	Authorship type			Total
	Nigeria only	Mixed	foreign	
Research data not available	41 (100%)	14 (87.5%)	38 (80.9%)	94(89.5%)
Research data available	0(0%)	2 (12.5%)	9 (19.1%)	11 (10.5%)
<b>Total</b>	<b>41</b>	<b>16</b>	<b>47</b>	<b>105</b>
Not applicable	6	0	4	10

Only 105 articles were included in the computer codes search. Table 3 shows that papers that were authored by Nigerian authors did not publish NLP computer codes, while 5.9% of the papers by mixed authors did publish the computer codes and 10.4% of papers that were authored by foreign authors published their computer codes.

*Table 3: Digitally archived computer codes by Nigerian and non-Nigerian authors*

	Authorship type			Total
	Nigeria only	Mixed	Foreign	
Computer codes not available	41 (100%)	15(93.75%)	43(89.6%)	102(94.3%)
Computer codes available	0 (0%)	1(6.25%)	5(10.4%)	6(5.7%)
<b>Total</b>	<b>41</b>	<b>16</b>	<b>48</b>	<b>106</b>
Not applicable	6	0	3	9

## 4. Conclusion and Recommendation

This study provides an insight into how researchers from Nigeria harness publication (open access) and research data (open data) sharing models of scholarly communication on digital archives, different from their counterparts from other parts of the world. The sample in this study is very small and not at all representative of publications from Nigeria and therefore the results cannot be generalized. Studies like this will require bigger bibliographic data before robust conclusions can be made on the open scholarship or digital archiving practices of researchers from Nigeria.

However, certain conclusions can be made. Though this study shows that publication sharing habit among Nigerian authors is strong, research data sharing on digital archive seems unpopular. It is

important that academic librarians create awareness on the importance and availability of non-commercial and institutional repositories for publication and research data sharing.

## References

- Asubiaro, T. (2017). An Assessment of the Cyber Presence of Academic Libraries in Nigeria. *African Journal of Library, Archives and Information Science*, 27(1), 65–76.
- Asubiaro, T. (2019). How Collaboration Type, Publication Place, Funding and Author's role affect Citations Received by Publications from Africa? A Bibliometric study of LIS research from 1996 to 2015. *Scientometrics*. <https://doi.org/10.1007/s11192-019-03157-1>
- Asubiaro, T. (2020). Evaluating the Availability and Openness of Resources in Nigerian Languages' Natural Language Processing Publications: Bibliometric and Content Analysis Approaches. *Canadian Journal of Information and Library Science*.
- Asubiaro, T., & Igwe, E. (2020). A systematized review of Natural Language Processing Research of Nigerian Languages. In A. Etim (Ed.), *Developing Countries and Technology Inclusion in the 21st Century Information Society*. IGI Global.
- Bakare, M. O., Munir, K. M., & Bello-Mojeeed, M. A. (2014). Public health and research funding for childhood neurodevelopmental disorders in Sub-Saharan Africa: A time to balance priorities. *Healthcare in Low-Resource Settings*, 2(1). <https://doi.org/10.4081/hls.2014.1559>
- Kari, K. H., & Baro, E. E. (2016). Digital Preservation Practices in University Libraries: A Survey of Institutional Repositories in Nigeria. *Preservation, Digital Technology & Culture*, 45(3), 134–144. <https://doi.org/10.1515/pdte-2016-0006>
- McKiernan, E. C. (2017). Imagining the “open” university: Sharing scholarship to improve research and education. *PLoS Biology*, 15(10), 25. <https://doi.org/10.1371/journal.pbio.1002614>
- Pinto, G. (2019). Training Software Engineers Using Open-Source Software: The Students' Perspective. 2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering Education and Training (ICSE-SEET), 147–157.
- Pouris, A., & Ho, Y.-S. (2014). Research emphasis and collaboration in Africa. *Scientometrics*, 98(3), 2169–2184. <https://doi.org/10.1007/s11192-013-1156-8>
- Raju, R., Adam, A., & Powell, C. (2015). Promoting Open Scholarship in Africa: Benefits and Best Library Practices. *Library Trends*, 64(1), 136–160. <https://doi.org/10.1353/lib.2015.0036>
- Tennant, J., Beamer, J. E., Bosman, J., Brembs, B., Chung, N. C., Clement, G., Crick, T., Dugan, J., Dunning, A., Eccles, D., Enkhbayar, A., Graziotin, D., Harding, R., Havemann, J., Katz, D. S., Khanal, K., Kjaer, J. N., Koder, T., Macklin, P., ... Turner, A. (2019). Foundations for Open Scholarship Strategy Development [Preprint]. BITSS. <https://doi.org/10.31222/osf.io/b4v8p>