

Institutional Priorities

CASE STUDY

Raising the global visibility of non-English, non-STEM research: a case study from Colombia



ELSEVIER

“The university’s 300-year tradition of social sciences and humanities research, shared in local languages, was largely invisible to the rest of the world”

Executive summary

Like many institutions of higher education, the University of Rosario (Universidad del Rosario) in Bogotá, Colombia is eager to maximize the visibility of its research capabilities and encourage international collaborations, especially in the social sciences and humanities. To this end, the university is employing three technological tools—the Pure research management system, the Mendeley reference management system and the DeepL translation service—to ensure that at least the titles and abstracts of all its non-English-language research are visible to and indexed in the major Internet search engines. The goal is to enhance the rankings of the university and its researchers and to facilitate learning and collaborations around topics of global and local concerns.

Introduction

A recent review of two large social sciences databases¹ revealed that nearly 90% of the contents are English-language publications. The increase in non-English-speaker users of the Internet and library databases “does not seem to have affected the language of scholarly publications in the social sciences,” where English continues to be the common language, according to author Frans Albarillo, a business librarian and professor at the Brooklyn College Library in New York City.

Further, UNESCO’s World Science Report 2010² showed that four countries—the United States, the UK, the Netherlands and Germany—together produced two-thirds of the social science journals registered in the largest databases of such journals. Asia contributed 9%, while Oceania, Latin America and Africa each contributed less than 5%.

Why do these statistics matter?

For one thing, the status of the institution is at stake. In correspondence published several years ago in *Nature*,³ scientists at Leiden University wrote that “the language of publication has a dramatic and largely underestimated effect on citation-based measurements of research performance.” Non-English-language journals generally have a low impact, as they have fewer readers. They also have fewer citations. Thus, in a system in which rankings are based on the number of citations per paper and per faculty member, “the language of publication directly affects the ranking of the university hosting the research.”

Of course, the individual researcher is also impacted. The H-index is a metric that measures both the productivity and citation impact of a scientist or scholar’s publications. The index is based on the author’s most cited papers, and the number of citations he or she received in other publications.

Similarly, Elsevier mines its Pure information management system to create an index of weighted terms that defines the key topic areas of an individual’s research.

In both cases, accuracy relies on having access to an individual’s entire body of research—not just those in English-language publications.

Ignoring non-English, non-STEM publications has additional consequences. Irina Bokova, former Director-General of UNESCO, wrote in the forward to the World Social Science Report 2010²:

“Social scientific endeavour is also poorer for its bias towards English and English-speaking developed countries. This is a missed opportunity to explore perspectives and paradigms that are embedded in other cultural and linguistic traditions.

A more culturally and linguistically diverse approach by the social sciences would be of tremendous value to organizations such as UNESCO in our efforts to foster mutual understanding and intercultural dialogue.”

The ability to access non-English publications also enables researchers and policymakers to place local phenomena in a global context, so that nations can learn from each other’s real-world experiences. And that’s where this case study starts.





Debunking narco fiction

In 2017, Carlos Estévez-Bretón, a former technology officer at the University of Rosario in Colombia, was at a conference when he was asked—not for the first time—about the accuracy of narco books and movies that take place in Colombia. He said that at least 70% was made up, but that people would never know because all the information covering what really happened was published in Spanish, and not in indexed journals. The research on narco-traffic, drugs, violence and other aspects of social conflict were in local publications because the researchers simply didn't work in English.

To deal with the aftereffects of the narco experience, global peacemakers looked to Ireland and South Africa, both of which had undergone long-term civil unrest, for solutions that might help Colombia. But any lessons learned from the application of those processes is in Spanish, meaning it would be difficult for the Colombian experience to help nations outside of Latin America.

Thinking about this issue was a turning point for Estévez-Bretón. He decided to undertake the challenge of making the university's non-STEM research more visible to search engines and, thereby, to the international community.

Surfacing social science research

Estévez-Bretón knew that visibility was not a problem for the university's researchers in medicine and the physical sciences, most of whom were accustomed to working in English or with English-speaking collaborators. Their work makes up about 60% to 70% of the university's research output.

But the university's 300-year tradition of social sciences and humanities research, shared in local languages, was largely invisible to the rest of the world.

He started with three tools: the Pure research management system, which includes a collaboration module; the Mendeley reference management system; and DeepL, a recently launched AI-driven translation service. The idea was to translate at least the title and abstract of all the Spanish social sciences publications into English.

Estévez-Bretón and his team began using a translation interface powered by DeepL to assist non-English-speaking researchers, who could then enter the English title and abstract as well as the original Spanish into the Pure research entry form. Pure would then serve up the English information to Google and other search engines.



Working through challenges

Currently, more than 1,000 publications need to be translated, at least in part, from the repository of existing social sciences research. And, of course, research is ongoing, so publications will be added continuously over time. Estévez-Bretón and his team tried several approaches to help gain momentum.

The members of the Research and Innovation Office handle special issues of research publications themselves. At the same time, they're informing professors that they will have to invest in translation and entry in their own departments in order to benefit from the system.

Not unexpectedly, there was a good deal of resistance, which Estévez-Bretón dealt with by explaining to the researchers that information systems are made up of three parts. The first two, software and hardware, are easy; they involve simply paying a vendor.

But the third part is people, and, generally, they are not included in the budget. Involving people to do the translations, enter the information into the system and keep the system up to date to reflect the latest research is an “educational and cultural” problem, he says.

First, researchers need to be made aware of two major results of doing their part: gaining visibility for the individual researcher

in the international community, and improving the visibility and ranking of the university and its research output.

Translating titles and abstracts into English and data entry need to become part of the social sciences and humanities routine. Many of the STEM indexed journals already require English titles and abstracts for papers published in Spanish and other European languages. But it's more complicated in the social sciences and humanities, where researchers are more likely to produce a book or book chapter. Mendeley helps by extracting information directly from these source files and generating bibliographic reference files that can be uploaded directly to Pure.

Although the university has not hired additional staff to support the initiative, several student interns are now recruited and trained each semester, providing about 60 hours each of translation and data entry time.

On the cultural side, many researchers who publish in local languages feel it doesn't make sense to try to translate their work into English because they wouldn't be able to capture the nuances of what they've done. But the result, Estévez-Bretón says, is that a good deal of knowledge, communicated not only in Spanish but in ethnic dialects, is being lost.

The problem is similar for other countries, such as Russia⁴ and the Arab nations. In fact, recent research shows that many languages are themselves in danger of being lost—and with them, the knowledge that they might provide.⁵

Currently, more than 1,000 publications need to be translated, at least in part, from the university's social sciences research.



Looking ahead

The drive to save dying languages is spawning International Organization for Standardization (ISO) standards for some of these languages, which is enabling people to configure their keyboards in non-mainstream languages, such as Cherokee, spoken by native Americans, and Quechua, a Latin-American indigenous dialect. Estévez-Bretón envisions a potential marriage between such standards and technology systems that could enable the translation of abstracts from various dialects, facilitating visibility of the work and providing opportunities for other researchers to make contact with the author.

Specifically for the University of Rosario, the technology team is looking first and foremost for changes in the Fingerprint® index of the university and of individual researchers. Powered by the Elsevier Fingerprint Engine®, the university's Fingerprint index is a concept graph created by mining the texts in Pure to categorize research topics and visualizing them on the university's research home page. Currently six topics are featured, all in the medicine and life sciences category. The team hopes to see topics in the social sciences or humanities featured in the near future.

Similarly, they hope that individual researchers will see and appreciate the emergence and expansion of their own Fingerprint graphs in their key subject areas—e.g., geopolitics, international relations, conflict. The system has been in place for only about a year, but that recognition is beginning to happen.

Conclusion

The primary goal of all these efforts is now in sight—namely, greater visibility in major search engines for all of the University of Rosario's research. That growing visibility is occurring largely because of the research portal, which is well indexed in Google and other search engines, and is also accessible to all Pure users globally, thus facilitating the identification of external experts, reviewers and potential collaborators.

References

1. Albarillo F (2014). Language in Social Science Databases: English Versus Non-English Articles in JSTOR and SCOPUS. *Behavioral & Social Sciences Librarian*, 33:77–90. DOI: 10.1080/01639269.2014.904693
2. UNESCO (2010). World Social Science Report. <http://unesdoc.unesco.org/images/0018/001883/188333e.pdf>
3. van Raan T, et al. (2011). Non-English papers decrease rankings. *Nature* 469, 34. <https://www.nature.com/articles/469034a>
4. Falkovich Y and Ostvald R (April 5, 2017). How we're making our university's research more visible. Elsevier Connect. <https://www.elsevier.com/connect/how-were-making-our-universitys-research-more-visible>
5. Nuwer R (June 6, 2014). Languages: Why we must save dying tongues. BBC. <http://www.bbc.com/future/story/20140606-why-we-must-save-dying-languages>

Resources

DeepL
<https://www.deepl.com/home>

Mendeley
<https://www.mendeley.com/>

Pure
<https://www.elsevier.com/solutions/pure>

University of Rosario research portal
<https://pure.urosario.edu.co/en/>

The Pure Community Module

The Pure Community Module facilitates the management of multi-institutional research collaborations, including data collation, analytics and shared reporting. It also provides a multi-institutional portal that enables the collation and aggregation of data to showcase the project's research assets, including the contributions of each collaborating institution.

Acknowledgment

The content of this paper and the figures contained therein have been reviewed by the Elsevier Librarian Audience team.

Contributing author:

Carlos Estévez-Bretón, former Technology Transfer Officer, University of Rosario, Bogotá, Colombia

Customer Consultant, Elsevier

Pure

As a highly versatile research information management system, Pure facilitates an evidence-based approach to your institution's research and collaboration strategies, assessment exercises and day-to-day business decisions.

For more information about Pure,
visit elsevier.com/pure

Elsevier offices

ASIA AND AUSTRALIA

Tel: + 65 6349 0222

JAPAN

Tel: + 81 3 5561 5034

KOREA AND TAIWAN

Tel: +82 2 6714 3000

EUROPE, MIDDLE EAST AND AFRICA

Tel: +31 20 485 3767

NORTH AMERICA, CENTRAL AMERICA AND CANADA

Tel: +1 888 615 4500

SOUTH AMERICA

Tel: +55 21 3970 9300

