

ACADEMIA | Letters

A reflection on the journey to build the first national science databases

Hoang Vuong

Abstract

How a senior researcher from a developing country can build an organic academic enterprise. Drawing from childhood experience with nature, past works with the business sector, and philosophy of data-driven research, the essay presents a compelling case of letting young graduates work on big database-building projects: one on Vietnamese social sciences; the other is more than 80 years of the pioneer science in Vietnam—mathematics. Two national databases have enabled meaningful data-driven interactions with scientific policymakers.

Keywords: Data-driven research projects; research team; interdisciplinary research

1. Introduction

Those were the mid-summer nights from my childhood, in the late 1970s, when the buzzing sound of cicadas overpowered the sweaty air, and we children, mosquito nets in hands, set out for our scientific inquiry. We would roam around late evening in search of juvenile cicadas and, in utter silence, holding our breath, watch how the insect wiggles and sheds its outer shell before spreading its transparent wings and taking flights. The enchanting moment was filled with our noisy debate about what and why it happened. Every time, the curiosity in me would soon be overtaken by an inexplicable sentiment—these cicadas are, in fact, so fragile and short-lived that they would all be gone after a few weeks of singing.

Academia Letters, June 2021

©2021 by the author — Open Access — Distributed under CC BY 4.0

Corresponding Author: Hoang Vuong, hoang.vuongquan@phenikaa-uni.edu.vn

Citation: Vuong, H. (2021). A reflection on the journey to build the first national science databases. *Academia Letters*, Article 1329.

From a young age, I have long been interested in observing natural phenomena and recording bits and bits of data, whether it be the number of metamorphoses we saw a night or the time it took for one to completely transform and fly away. The memories have stayed with me as I took on different career paths, all of which are rooted in curiosity to explain today's data-centric world. Since I published my first academic paper in 1997, I have done work for universities, governments, and industries and started my own businesses. Much of my main work has involved data preparation, database design and construction, data-processing applications, and data analysis. These experiences have allowed me to observe the remarkable parallels in the seemingly "impossible triangle" of nature, academia, and entrepreneurship.

In the middle of this impossible triangle, data rationalizes and weaves nature, academia, and entrepreneurship coherently. Even though I did not have any idea during those mid-summer nights, the patient observation of cicadas is a form of data collection. While the data that I collected back then was used for noisy debates with friends, in a way, they prepared me for my long career in business and academia. In 2017, I founded a database named SSHPA to systematically analyze Vietnamese social scientists' research capacity since 2008 to aid with science policymaking (Vuong et al., 2018). Later, the project was funded by Vietnam's National Foundation for Science and Technology Development (NAFOSTED). Later, I co-founded another database that covers 80 years of mathematic development in Vietnam, with its code name being "SciMath DB," with Prof. Ngô Bảo Châu (a 2010 Fields Medalist), the Vietnamese mathematician of the University of Chicago—awarded the top mathematics honor together with the French mathematician Cédric Villani. Prof. Ngô Bảo Châu initiated the SciMath DB Project in the role of Scientific Director of the Vietnam Institute for Advanced Study in Mathematics (VIASM).

Throughout these data-driven projects, I have been able to foster an interdisciplinary research team that embraces the beauty of nature, the decisiveness of business, and the curiosity of science.

In what follows, I am trying to re-describe the journey of how an interdisciplinary research team was established and then how social data in a developing country have fueled scientific exploration.

2. Scavenging recourses, with the right timing and serendipity

Both academics and entrepreneurs will find that successes and failures are often defined by timing and serendipity (Vuong, 2016). Having the right mindset to catch and nurture those serendipitous moments and ideas is the key. For example, in late 2016, seeing that science policymakers were at a crossroad of revamping domestic research capacity in line with in-

ternational standards, I realized that, with my limited resources, this was the only window of possibility to build the SSHPA database (Vuong et al., 2018). The primary reasons were the small number of Vietnamese authors with international publications at the time, which made it feasible to track and verify each record, and the potential for commercialization. Having confidence in one's ideas and seizing the opportunity enables us to maximize our chance for career advancement and social impacts. Eventually, the SSHPA database also impacted the Vietnam Ministry of Education and Training's Circular 08, which focuses on requiring Ph.D. candidates, and their supervisors to publish their works internationally in 2017 (Nguyen et al., 2019). Such a meaningful data-policy interaction has encouraged me to continue the journey.

Both academics and entrepreneurs often start with very little resources; thus, managing and growing limited resources is a key question of survival (Vuong, 2018). The most important resource, for me, are ideas and mindsets. Finding the right ideas and having the right mindsets allow you to attract the right people and grow your organization *organically*. Thus, in 2017, I assembled a team of four young graduates for the SSHPA database (Vuong et al., 2018). I trained the team members from the baby steps of data identification, collection, storage, organization, categorization to the more complicated tasks of software engineering support and database system management. With our core values defined, early investment in human capital proves to be vital for the long-term sustainability of the academic enterprise.

Letting the young graduates face the endlessness of data in databases such as SSHPA or SciMath is mental preparation for scientific research's hardship. They offer a unique insight into the nature of science, especially social sciences. Some problems need to be addressed. Some thoughts need to be jotted down. There are reviews that need to be rebutted. It resonates deeply with the image of Sisyphus that Albert Camus used to conceptualize the absurdity (Camus, 2013). As his eternal punishment, Sisyphus had to push a rock up a mountain so it can roll back down again. For eternity, he does it over and over again. Through interaction with data, the young graduates can confront the endlessness, embrace it, continuously learn and unlearn, and find happiness from such an absurd condition (Vuong, 2016). This courage of conviction is very important to find like-minded colleagues and build a culture for a research team.

3. Courage and perseverance to walk the dark road

It might be a cliché, but both academics and entrepreneurs find the most success when they have the courage to walk the untrodden path, climb the mountains, and go the extra miles. When Ngô Bảo Châu and I decided in 2017 to build the SciMath database, we did not have any funding. All we had was the glorious tradition and history of Vietnam mathematics with

keen interest and engagement from well-known figures such as Laurent Schwartz (1915-2002; 1950 Fields Medalist) and Alexander Grothendieck (1928-2014; 1966 Fields Medalist), or Hoàng Tụy's (1927-2019; 2011 Constantin Carathéodory Prize) internationally recognized pioneering work on global optimization, which brought him the honor of the "Tuy cuts" (Tuy, 1968) and the "DC Programming" (Tuy, 1986). We decided to follow our entrepreneurial instinct and self-funded the project while waiting for the right investor or business partner to come.

Three years in, not only have we secured sufficient funding from the private sector, but we have also completed collecting electronic records of more than 10,000 indexed mathematic papers (co-)written by Vietnamese authors (Chau et al., 2020). To be exact, as of March 10, 2021, the SciMath database has recorded 10,475 papers, 74 years, and 1,995 mathematicians. The data count can feel lifeless, but they were contributed by papers written in the times of bloody wars from the 1940s to the 1970s, political conflicts, economic and communication barriers (Thiem, 1949, 1950; Tuy, 1968). Figure 2 below demonstrates the network of Vietnamese mathematicians from 2010 until 2021:

Given that mathematics is seen as a pioneer in the sciences in Vietnam (Nguyen et al., 2021), tracing its development for nearly a century allows us to understand history better. It reminds us of numerous generations of Vietnamese mathematicians who benefited from the works of intellectual architects. Many have become influential scholars in their fields after working in the Institute of Mathematics, founded by Hoàng Tụy.

In this sense, I see the project very much like a banyan tree, which takes years to set its deep roots and grow out its massive trunk and branches. The data walks us through the works of many intellectual architects, how they set the roots of Vietnamese mathematics. Hoàng Tụy and Lê Văn Thiêm, for instance, initiated the Institute of Mathematics, the cradle for the career of many influential Vietnamese mathematicians. We learned how Ngô Bảo Châu's recipient of the Fields medals set a new milestone in mathematical research, how the event gave people the courage to climb the academic ladder. Figure 3 presented the number of mathematic articles ten years before Ngô Bảo Châu's achievement (2000-2010) and ten years after (2011-2020). From 2000 until 2010, only 2531 articles were recorded. The number of mathematics articles doubled in the next 10-year period with 4474 articles. The achievement of Ngô Bảo Châu also allowed enormous resources to be invested in mathematics. The most notable being the National Key Program on Mathematics Development in the 2011-2020 period (Vietnam Government, 2010). The Prime Minister approved the program under Decision No. 1483/QĐ-TTg, with the funding of VND 651 billion (approximately USD 28.5 million).

Besides providing an overview of the mathematical research in Vietnam, the SciMath database can also help policymakers identify research gaps, rising mathematicians, or strong research centers. For instance, the SciMath database has identified 23 out of 1995 Vietnamese authors with publications in top mathematics journals (see Table 1). The information will be crucial for policymakers in the future, just as in the case of the SSHPA database and Circular 08.

Author	Number of articles	Number of top journals
Pham Huu Tiep	21	6
Vu Ha Van	17	5
Duong Hong Phong	13	6
Dinh Tien Cuong	6	5
Ngo Bao Chau	6	4
Ngo Viet Trung	6	3
Lu Hoang Chinh	5	4
Nguyen Hoai Minh	4	4
Nguyen Trong Toan	3	2
Nguyen Huu Hoi	2	2
Nguyen Sum	2	1
Nguyen Duy Tan	2	1
Pham Hoang Hiep	2	2
Pham Hung Quy	2	1
Pham Ngoc Anh	1	1
Phung Ho Hai	1	1
Nguyen Huu Anh	1	1
Nguyen Tien Dung	1	1
Nguyen Van Hoang	1	1
Nguyen Cong Minh	1	1
Nguyen Dang Hop	1	1
Dao Hai Long	1	1
Ha Huy Khoai	1	1
Le Van Thiem	1	1

Table 1. Mathematics authors with articles in top mathematics journals

The dataset, insights, and remaining unanswered questions have contributed to the national policymakers' decisions to continue government funding to the project and reaffirm support to VIASM in the next ten years.

4. Risk management and goal setting

To survive and thrive, both an academic and an entrepreneur must learn to manage risks, set goals, deal with inevitable setbacks in their journeys. In my experience of building two large databases, understanding the different levels of risk involved in each project—with 80 years of data for the SciMath project posing a more daunting prospect, I learned the importance of choosing a challenge that matches our potential at the time. For this reason, we kicked off the SSHPA project first in 2017, ensuring that our system could handle ten years' worth of data before embarking on the more ambitious project of SciMath in early 2020. To continually test one's growing capacity contributes to cultivating a sustainable research culture.

Moreover, as the young graduates grow with the process of data collection, it is important for them to reflect on their own journeys. Certainly, while big projects like SSHPA or SciMath can be a project for veteran researchers, it can mean a lot for young researchers. Working on the projects amid the reproducibility crisis (Munafò et al., 2017), the young graduates can deeply understand the importance of data. The trickiness of a situation when the findings will never be corrected unless the authors or someone else realize that something is off cannot be taught easily without working with data firsthand (Haas, 2021).

5. Conclusion

In a world where even basic facts get repeatedly questioned (Kata, 2010), I often think back about how I have arrived at countless moments of truth by observing nature and curiosity in science. The transient nature of life is perhaps something we all think about from time to time. As I relive the summer nights and the cicada metamorphosis at fifty years old, I also reflect on my journey as an academic entrepreneur. Like a pendulum, the journey finds its equilibrium in asking questions and finding answers through collecting data and using an appropriate scientific method.

Building an academic enterprise, as I have learned, means finding the elusive answers in the deluge of data and social problems today. This is all the more relevant as uncertainty from the COVID-19 pandemic continues to envelop our world. I think social scientists can create a variety of products and services from the raw social data streams. In doing so, we bring our expertise to bear on the great task of finding much-needed social innovations for our social structures, which have been subjected to the break-neck speed of computational development.

Last but not least, as a senior researcher, I also think about paying it forward. Hopefully, the kind of academic enterprise that we build can ease some of the struggles faced by early-career researchers, sowing the seeds for a healthy academic ecosystem for the next generations.

References

- Camus, A. (2013). *The myth of Sisyphus*: Penguin UK (Originally published in 1942).
- Chau, N. B., Hoang, V. Q., et al. (2020). The 80-year development of Vietnam mathematical research: Preliminary insights from the SciMath database on mathematicians, their works and their networks. *Technical Report No. VIASM-AISDL-20.02, presented at VIASM Scientific Council Meeting on November 13, 2020*. Retrieved from DOI: arXiv:2011.09328 (March 10, 2021)
- Haas, B. D. (2021). What my retraction taught me. *Nature*, 589, 331. doi:https://doi.org/10.1038/d41586-021-00073-4
- Iborra, L. (n.d.). Penrose Triangle. *The Noun Project*. Retrieved from https://thenounproject.com/search/?creator=2129742&q=Penrose&i=881121 (March 10, 2021)
- Kata, A. (2010). A postmodern Pandora's box: Anti-vaccination misinformation on the Internet. *Vaccine*, 28(7), 1709-1716. doi:https://doi.org/10.1016/j.vaccine.2009.12.022
- Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., . . . Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, 1(1), 0021. doi:10.1038/s41562-016-0021
- Nguyen, H.-K. T., Nguyen, T.-H. T., Ho, M.-T., Ho, M.-T., & Vuong, Q.-H. (2019). Scientific publishing: the point of no return. In Q. H. Vuong & T. Trung (Eds.), *The Vietnamese Social Sciences at a Fork in the Road* (pp. 143-162). Warsaw, Poland: De Gruyter.
- Nguyen, T. T. H., Pham, H.-H., Vuong, Q.-H., Cao, Q.-T., Dinh, V.-H., & Nguyen, D. D. (2021). The adoption of international publishing within Vietnamese academia from 1986 to 2020: A review. *Learned Publishing*, 33(4), EarlyView. doi:https://doi.org/10.1002/leap.1340
- Thiem, L.-V. (1949). Über das Umkehrproblem der Wertverteilungslehre. *Commentarii Mathematici Helvetici*, 23(1), 26-49. doi:10.1007/BF02565590
- Thiem, L.-V. (1950). Sur un problème d'inversion dans la théorie des fonctions méromorphes. *Annales scientifiques de l'É.N.S.*, 3e(67), 51-98.
- Tuy, H. (1968). On linear inequalities. (English. Russian original). *Soviet Mathematics. Doklady*, 9, 366-369.

- Tuy, H. (1986). A General Deterministic Approach to Global Optimization VIA D.C. Programming. In J. B. Hiriart-Urruty (Ed.), *North-Holland Mathematics Studies* (Vol. 129, pp. 273-303): North-Holland.
- Vietnam Government. (2010). Decision No. 1483/QĐ-TTg dated August 17, 2010, of the Prime Minister approving the national key program on mathematics development in the 2010-2020 period. *Viet Nam Government Portal*. <https://bit.ly/3xI93ut>
- Vuong, Q.-H. (2018). The (ir)rational consideration of the cost of science in transition economies. *Nature Human Behaviour*, 2(1), 5-5. doi:10.1038/s41562-017-0281-4
- Vuong, Q.-H., et al. (2018). An open database of productivity in Vietnam's social sciences and humanities for public use. *Scientific Data*, 5(1), 180188. doi:10.1038/sdata.2018.188
- Vuong, Q. H. (2016). Global mindset as the integration of emerging socio-cultural values through mindsponge processes: A transition economy perspective. In J. Kuada (Ed.), *Global Mindsets: Exploration and Perspectives* (pp. 109-126). London: Routledge.