

Reflections on a Quest for Social Science Data in Vietnam

As a developing country likely to enter the World Trade Organization in a few years, Vietnam is, perhaps not surprisingly, awash in social science studies, as government mandates, donor-funded projects, and new research opportunities contribute to a renewed focus on a country seeking very hard to enter the global mainstream. Indeed, as I suggested in my Fulbright research proposal or “project statement” drafted in 2002:

As Vietnam moves toward entry into the World Trade Organization, and its economic system becomes more mixed, social science research data can be expected to play an increasingly important role for scholars interested in understanding and analyzing Vietnamese society, as well as for policymakers within Vietnam and those in the United States [and I might add here, elsewhere.]. Scholars and their students will also be interested in using such data for secondary analysis.

Vietnam has already embraced the Internet. The country has an educated, motivated, and increasingly urbanized sector of the population, although it remains basically an agricultural society. But even in some rural areas, and certainly in urban enclaves like Ho Chi Minh City and Hanoi, Internet cafes have proliferated in just a few years, many with high-speed ADSL connections.

Donor agencies have for the last several decades poured millions of dollars into Vietnam. With the American embargo lifted a decade ago, the U.S. has now become the biggest trading partner with Vietnam, with more goods exported to the U.S. than vice versa.

Cable or satellite television has penetrated not just the expatriate community but also working-class neighborhoods. In fact, in my own apartment, while flipping the 16 channels on my TV one day, I saw the Dalai Lama in a news clip speaking to students at my university while wearing a cap with the school initials, UCI. That was on CNN; in addition to three Vietnamese-language channels (one of them aimed at overseas Vietnamese), others included CCTV from China, Deutsch World, and TV5 (France).

by Daniel C. Tsang¹ *

I mentioned that Vietnam is “awash in social science studies.” One indication of that is the prolific output from the country’s General Statistics Office (GSO), which publishes many times a year thick compendia of socio-economic statistics, gathered from national surveys and administrative agencies.

Annually, it publishes a bilingual (Vietnamese/English) *Statistical Yearbook*, now available in a CD-ROM edition as well. One compendium I was able to pick up was *Major Socio-economic Information Obtained from Ten Large Scale Surveys in Period 1998-2000*. Hanoi: Statistical Publishing House, 2001. It was a heavy volume; it ran over 2,000 pages and included data from not only the ten different surveys but also “main socio-economic indicators of 61 provinces and cities.” The chapter headings for the ten surveys, taken verbatim, give some idea of the proliferation of survey data in just three years in Vietnam:

1. Land census 2000
2. Population and housing census 1/4/1999
3. Labour and job survey 1/7 1998, 1999, 2000
4. Development investment capital survey 2000
5. Results of inventory forest 1999
6. Industrial complete survey 30/6/1998
7. Survey on state-owned trade and services companies 30/6/1998
8. Survey on non-state trade and services companies 2000
9. Poverty survey 1999
10. Survey on infrastructure of communes in rural areas 1998, 1999, 2000

An earlier volume covered 1990-1996.

Most importantly, the GSO in December, 2004 published a bilingual (Vietnamese/English) and massive 4,856-page, three-volume compendium, *Vietnam Statistical data in the 20th Century*. The first two volumes covered socio-

economic data from 1901-2000. Volume 3 featured results of 21 large-scale surveys and censuses.²

In addition to data in tabular format, reports from GSO-conducted surveys have been made available in print and, occasionally, online -- for example, *Vietnam: Demographic and Health Survey, 2002*. While the GSO sells the print volume of the report, its electronic PDF version is downloadable for free from the MEASURE DHS+ Demographic and Health Surveys Web site [URL: http://www.measuredhs.com/countries/country.cfm?ctry_id=56], which also makes the 1997 survey report available.

Similarly, the report from the well-respected Vietnam Living Standards Survey 1997-1998 is also downloadable for free from the Web, despite the fact that the GSO sells the paper edition in bookstores in Vietnam. The URL for the report and associated documentation is: <http://www.worldbank.org/html/prdph/lsms/country/vn98/vn98docs.html>. The report from an earlier living standards survey, that of 1992-93, is also available with related documentation on the Web at: <http://www.worldbank.org/html/prdph/lsms/country/vn93/vn93docs.html>.

Unfortunately, the datasets themselves from which these reports are derived are not available online for free distribution. They can be acquired, for a fee, from the GSO itself. A graduated fee structure is detailed on the World Bank living standards site, depending on whether the requester is a Vietnamese or foreign organization or individual and whether they are located in a developing country or not. Fees range from \$200 a dataset to \$2,000.

The conditions for use include the unremarkable (e.g., "The user must respect the copyright of the data and should not allow other users to use the data without permission of the GSO") but also the potentially problematic: "The user is requested to send to the GSO the results of use of the data at least one month before their publication."³

Although not commenting directly on the Vietnam data, but rather on data access to all living standards surveys, the World Bank notes on its Web site⁴ that "According to our best information, a substantial proportion of data requests have been denied, left unanswered, or answered affirmatively only after substantial delays, or there is not yet an established track record."

That is not to say that researchers seeking to acquire all GSO datasets have to go through numerous hurdles. In a striking example of data sharing that bodes well for the future, a GSO team⁵ worked with historian Bob McCaa at the University of Minnesota to make the 1989 and 1999 Vietnam census available in 2002. As a result, the two latest decennial censuses of Vietnam are available via request to IPUMS International [URL: <http://www.ipums.umn.edu/international/index.shtml>]. The survey instruments are

available here: http://www.ipums.umn.edu/international/source_materials.shtml.

In addition, the GSO is the lead agency working with other government and non-governmental data producers in Vietnam in an exciting new venture, gathering and contributing data on a host of socio-economic indicators to the UN Development Program's DevInfo/VietInfo database, which will initially be disseminated on a free CD-ROM.⁶

The GSO has also improved its marketing, publishing a handy and colorful 22-page bilingual (Vietnamese/English) pamphlet, "Introduction of Main Statistics Products and Services of Viet Nam General Statistics Office" (Hanoi: General Statistics Office, 2003), listing its print and CD-ROM products. According to information from the pamphlet itself, 1,000 copies were printed. How it was distributed remains unknown to me. A new website provides, in Vietnamese and English, latest statistical developments as well as socio-economic data [URL: http://www.gso.gov.vn/default_en.aspx?tabid=491].

I was able to obtain a free copy of that guide while managing to locate, on GSO premises, its unheralded newly established "GSO data center," where researchers as



well as members of the public can obtain tabular socio-economic data from the helpful staff, at least one of whom spoke English. It was also there that I was able to purchase the latest GSO publications and CD-ROMs.

Realistically speaking, many hurdles remain before data sharing can be a fully realized reality in Vietnam. A common refrain I heard from researchers is that "data is power," giving rise not to a culture that values sharing data for free or at reasonable cost, but on that seeks to sell it to the highest bidder. This is, of course, exacerbated by the fact that local salaries are quite low (a state employee

may earn around US\$40 in basic pay monthly). Thus, the temptation to supplement meager wages is ever-present.

In addition, donor agencies and government directives merely require the end product of a project to be the creation of a published report, with no attention paid to the datasets used to compile the reports.

Furthermore, the infrastructure for broadband connectivity to the Internet in Vietnam needs to be improved. The Vietnamese Academy for Social Sciences has yet to implement broadband access for all its component research institutes (let alone libraries), and in some universities, the server serving as the Internet gateway remains intolerably clogged like some California freeway at rush hour.

Also, it is unclear how stricter Internet regulation in Vietnam, as recently reported, will impact scholarly Web sites within Vietnam.

Finally, there is as yet no in-country, accessible data archive at either the national or the sub-national level that could serve as the location to store the data. Much of the library world attention in Vietnam still is on digitizing print material, if the digital library is at all part of the dialog, and not yet on data files.

However, other selected data and/or metadata (documentation) have made the migration to the Web, in addition to the GSO data already mentioned. These include University of Washington sociologist Charles Hirschman's Vietnam Living History Survey 1991 (URL: <http://www.csde.washington.edu/research/vietnam/vlhswebdocs/data.html>) and his Vietnam Longitudinal Survey 1995, conducted in collaboration with the Institute of Sociology in Hanoi. In addition, in 2001, my school collaborated with the Institute of Human Studies in Hanoi to conduct, for the first time, the Vietnam component of the World Values Survey [URL: <http://www.democ.uci.edu/democ/archive/vietnam.htm>]. Papers relating to that survey are available on that site, and the data are available from ICPSR (<http://webapp.icpsr.umich.edu/cocoon/ICPSR-STUDY/03975.xml>). Another Vietnam wave will be conducted in 2005.

I am indebted to Cristina Negrut at the Davidson Center at the University of Michigan for her U.S. National Science Foundation Web site that serves as a gateway to business and economic datasets and/or documentation relating to "transitional" or "emerging" economies, including Vietnam. I am in the process of creating a Web site dedicated to Vietnam social science data, and a preview of some of the resources to be included is available on the "Data Sources" page of the UCI Social Science Data Archives [URL: <http://data.lib.uci.edu>]. Scroll down to Vietnam or click on V in the alphabetical listing on the Data Sources page to reach the Vietnam data section.

Indeed, exciting data developments can be anticipated in Vietnam, which hosted a major GIS conference, "GIS Ideas 2004" [URL: <http://gisws.media.osaka-cu.ac.jp/gisideas04/>] in September 2004. The conference was the International Symposium on GeoInformatics for Spatial-Infrastructure Development in Earth & Allied Sciences, which includes among its goals the development of spatial databases and analysis of spatial data relating to natural and social environments. Press accounts of the gathering included this dispatch from the official Vietnam News Agency:

According to Dr Hoang Dung of HCM City's National University, while GIS projects have shown remarkable results, the Government is yet to create a detailed plan for its further application.

"Though the 70 GIS projects, at a total value of VND85 billion (US\$5.4 million), have achieved remarkable results, the Government and the Ministry of Science and Technology (MoST) have not created a detailed blue print for its further exploitation, either in localities or at the national level," he said.

"The MoST has to make a plan to develop a GIS data warehouse and link it with localities and international organisations," suggested Dr Tran Vinh Uoc of the National University, HCM City.⁷

Also, a Vietnam Economics Research Network [URL: <http://www.vern.org.vn/>] has been set up. Its Web site includes a link to the Hanoi-based Institute of Economics' Vietnam Trade Database [URL: http://www.vern.org.vn/TradeDatabase/About_Database_E.htm], where one can search [URL: <http://www.vern.org.vn/TradeDatabase/Search/searchE.asp>], import, and export data online. It was created with technical support from the GSO.

I don't want to end without saying this: Data sharing is not a one-way effort, not just for researchers in the rich countries to obtain data from the developing world. It needs to be a two-way road. Yet researchers I met in Vietnam are unable to afford data from the West. We in the data sharing and data archiving community must do better in not only technology and expertise transfer but also in seeking ways to establish the infrastructure necessary for making western data available to developing countries at greatly reduced costs, following the example of UNESCO in developing countries. Many print libraries I visited, for example, use library cataloging software distributed gratis by UNESCO.

For example, could the major U.S. data archive, ICPSR [URL: <http://www.icpsr.umich.edu>], consider a membership pricing scheme that encourages use from researchers, institutes, Nongovernmental Organizations (NGOs), and government agencies in a developing country

like Vietnam? I am told that ICPSR considers this on a “case by case” basis. Or in the realm of technology transfer, could licensing for the data publishing utility, Nesstar [URL: <http://www.nesstar.com>], be priced to make its software available more widely? I understand they are open to this.

In addition, it doesn't make sense for donor agencies to continue to pour money into producing surveys, just to end up with a printed report. The problems faced by researchers seeking to acquire the World Bank's Living Standards surveys have already been alluded to. Donor agencies must look beyond the paper product. It would make more sense to have granting agencies require, as a condition of a grant, deposit of the data and associated documentation, for example, in a data archive, even an emerging one, in-country or abroad, with the additional condition that the data be made shareable at minimal or no cost. A certain percentage of the grant can be slated for data cleaning, archiving, and dissemination. And why not devote another percentage toward a fund to set up an in-country data archive so that data are not “lost” but can be preserved and made shareable? It is unproductive for donor agencies to fund projects that “reinvent the wheel” each time.

We need to think creatively about what else we can do to make data sharing a more realistic option for our colleagues abroad, especially in Southeast Asia and other developing countries. Otherwise, we perpetuate a world of data haves and have-nots. That is the challenge I leave with you.



* Paper prepared for presentation at the IASSIST Annual Conference, University of Wisconsin - Madison, May 26 2004.

Footnotes

¹ The author is Social Science Data Librarian and bibliographer in Asian American Studies, Political Science,

Economics and Asian Studies (acting) at the University of California, Irvine. He did field work from January to early May 2004 as a Fulbright Research Scholar in Hanoi, Socialist Republic of Vietnam, where he was based at the Institute of Sociology, Vietnamese Academy of Social Sciences. He did follow-up work in August-September 2004 in Vietnam. With several dozen scholars and fellows (students) each year, the Fulbright program in Vietnam remains, remarkably, the best-funded Fulbright program in the world. The author is especially interested in hearing from others working on data sharing in Southeast Asia. Write him at: dtsang@uci.edu.

² *So Lieu Thong Ke Viet Nam The Ky XX = Vietnam Statistical Data in the 20th Century.*

Hanoi, Vietnam: Nha Xuat Ban Thong Ke, 2004. (See: http://www.gso.gov.vn/default_en.aspx?tabid=494&itemid=2235&idmid=1.)

³ “Distribution of the 1992/93 and 1997/98 Viet Nam Living Standards Surveys,” Living Standards Measurement Study of the World Bank [URL: <http://www.worldbank.org/html/prdph/lms/country/vn98/vndata.html>]. Accessed 5/21/04.

⁴ “Selecting an LSMS Data Set,” Living Standards Measurement Study of the World Bank [URL: <http://www.worldbank.org/html/prdph/lms/guide/select.html>]. Accessed 5/21/04.

⁵ “Vietnam Census Microdata Integration Project,” Integrated Public Use Microdata Series International [URL: <http://www.hist.umn.edu/%7Eermccaa/IPUMSI/vietnam/index.html>] Accessed 5/21/04.

⁶ Information from Nguyen Bich Ngoc, UNICEF Vietnam, in a meeting September 8, 2004, at UNICEF Vietnam in Hanoi, Vietnam.

⁷ “GIS technology benefiting VN,” *Viet Nam News*, September 25, 2004. Available online: <http://vietnamnews.vnagency.com.vn/2004-09/24/Stories/10.htm>.