

UC Irvine

LAUC-I and Library Staff Research

Title

Public Opinion Archiving and Data Sharing in Asia

Permalink

<https://escholarship.org/uc/item/7521m9s4>

Author

Tsang, Daniel C

Publication Date

2016-05-12

Public Opinion Archiving and Data Sharing in Asia

By Daniel C. Tsang¹

It may be self-evident the need for archiving survey data – including public opinion data, but it may be useful to reiterate the benefits of having such data preserved. Indeed, it is part and parcel of social science research and especially in Asia, with the growth of civil society and the concomitant need to have data that informs not just policy makers but also the citizenry so that all can become full participants in a transparent, civil society. As societies become more democratic, the voice of the citizenry becomes more vocal and more important, and public opinion surveys are one major way policy makers become aware of what can be at stake. In addition, as more and more students enter social science disciplines, the need for more accessible data (for secondary analysis) becomes much more important.

The challenges faced by data archiving of social science data in Asia, including public opinion surveys, are probably similar to manifestations elsewhere, but cultural and political factors are likely to play a much more decisive role here.

In the West, establishing data archives to further survey researchers' needs is a given. ICPSR, the leading social science data consortium, is already over four decades old, having celebrated its 40th anniversary in 2003. IASSIST, the data archivists' professional organization, in which I am actively involved, already celebrated its 30th anniversary last year. Roper Center, the premier public opinion data archive, is similarly well established. With the fall of communism in Eastern Europe, western agencies helped fund the founding of social science data archives throughout the region, so that there is one archive in Estonia, Bulgaria, Lithuania etc. Many of these archives have taken advantage of technological advances to implement increasingly user-friendly data access policies.

Both ICPSR and now Roper have implemented "direct downloading" of datasets for its membership – so that university campus users at any computer workstation can now have access to archived public opinion polls, among other studies.

The need to develop social science information sharing in Asia is not just a recent phenomenon. Back in the 1980s, UNESCO formed an Asia-Pacific Information

¹ Social Science Data Librarian at the University of California, Irvine, since 1986, managing its data archive (<http://data.lib.uci.edu>). He is also associated with the Center for Asian Studies at UCI. He serves on the administrative committee of the leading social science data archivists' professional organization, IASSIST (<http://www.iassistdata.org>). He is also a fellow of the International Professors Project (<http://www.internationalprofs.org>). While working at ICPSR's archiving unit in the 1970s, he merged the 1972, 1974 and 1976 American National Election Studies into one data file and processed the Miller, Miller and Kline's Media Content Analysis Study, 1974. In 2004, he spent several months in Vietnam conducting research into data archiving there as a Fulbright research scholar. This paper is one outcome of that Fulbright-funded research. It is also supported in part by a generous grant from the Center for Asian Studies at UCI. His public opinion subject guide is here: http://www.lib.uci.edu/online/subject/subpage.php?subject=public_opinion. Contact email: dtsang@uci.edu. This paper was first presented at the WAPOR regional meeting in Hong Kong, December 8-10, 2005.

Network in Social Sciences (APINSS), which issued a report² in 1986 of its regional meeting. That report is the earliest effort at strategizing the need for better preserving social science information in Asia. Equally significant, it was apparently the first time social scientists in the region got together to discuss how to share such information. Significantly, its membership – the group still continues operating -- includes both Koreas: The Democratic Republic of Korea's Academy of Social Sciences and the Republic of Korea's Korea Institute of Industry and Technology Information.³

Public opinion data of past few decades have always been digitized – and now increasingly are “born digital” – but distributing the data has had to await the maturation of the Web. This is now possible. Ironically, the major public opinion archive relating to an Asian country resides outside Japan: With more than 10,000 questions and responses, JPOLL is the “only comprehensive database of Japanese public opinion data” and is located at the University of Connecticut's Roper Center:

<http://www.ropercenter.uconn.edu/jpoll/JPOLL.html>. Funded by grants from the United States-Japan Foundation, it provides licensed access to members of the Roper Center consortium, including many universities.

Researchers have pointed out that in the emerging economies of Asia, there still remains a digital divide. South Korea may boast the highest Internet connectivity yet many places in Southeast Asia, for example, remain unconnected. For example, when I was teaching in Vietnam last year, the university had broadband connectivity in one sparkling clean classroom (used by “foreign guests”), but the connection was so slow that I gave up and just used screen dumps the next day. Yet Vietnam, in the prelude to its likely entry into the WTO, is already “awash in social science studies” with IGOs and NGOs funding projects that do many surveys of living conditions across the country.⁴ Accordingly only a few Asian countries have established in-country major social science survey data archives: Japan (the major one being Social Science Japan Data Archive:

<http://ssjda.iss.u-tokyo.ac.jp/en/>), South Korea, (Korean Social Science Data Center: <http://www.ksdc.re.kr/unisql/engjap/eindex.html>)⁵, Taiwan (Survey Research Data

² *Social and Human Sciences in Asia and the Pacific; Asia Pacific Information Network in Social Sciences (APINSS): Report of a Regional Meeting*. Bangkok: Unesco Regional Office for Education in Asia and the Pacific, 1986.

³ The network's most recent membership is listed here:

<http://www.riclib.nrct.go.th/link/epcountry.html>

⁴ Daniel C. Tsang, “Reflections on a Quest for Social Science Data in Vietnam,” *IASIST Quarterly* 28/1 (2004).

⁵ For its origins, see: Nitta Michio, “Center Court: An Introduction to the Information Center for Social Science Research on Japan,” *Social Science Japan* 10 (August 1997), 3-5; Matsui Hiroshi, “Plans for Data Archives at the Information Center for Social Science Research on Japan,” *Social Science Japan* 10 (August 1997), 9-10 and “Archiving Data for Social Scientists,” *Social Science Japan* 10 (August 1997), 11-13.

Archive ⁶, Philippines (Social Weather Stations: <http://www.sws.org.ph>) and Australia (Australian Social Science Data Archive: <http://assda.anu.edu.au/>).⁷

China, which now has a federated membership in ICPSR, is beginning to promote the availability of social science data. Its statistical agencies have worked with the China Data Center (<http://chinadatacenter.org/newcdc/>) at the University of Michigan to offer digitized provincial and regional statistical yearbooks as well as census information through the licensed database, China Data Online. The China Data Center (and this is separate from ICPSR) is also planning to provide access to Chinese surveys while within China itself, the China Population and Development Research Center (which holds the ICPSR membership) is now offering datasets to researchers. There is also the China Data Archive at Texas A&M (<http://chinaarchive.tamu.edu/portal/site/chinaarchive>) which includes various surveys as well as geospatial data among other datasets.

Cross-national surveys that include Asian countries have also made their data available on the Internet, among them:

Asia Barometer

<http://avatoli.ioc.u-tokyo.ac.jp/~asiabarometer/>

Based in Japan, this survey has conducted three waves covering various Asian countries.

East Asian Barometer

<http://eacsurvey.law.ntu.edu.tw/>

This Taiwan-based survey covers six countries, China, Taiwan, Hong Kong, the Philippines, South Korea and Thailand.

World Values Survey

<http://www.worldvaluessurvey.org/>

Surveys conducted now in over 60 countries, including many in Asia, since 1981.

International Social Survey Program

<http://www.issp.org/homepage.htm>

A survey of social conditions across 39 countries.

There is also the Korea Barometer Data Center (<http://www.koreabarometer.org/>) whose site provides contact information for free access to various Korea Barometers conducted from 1994-2001.

Barriers to data sharing

Data archiving presupposes a commitment to share data, yet data sharing faces many challenges. I'll just touch on a few.

⁶ Alfred Ko-wei Hu, "The Development of a Survey Data Archive in Taiwan," IASSIST conference, Madison, Wisconsin, 2003. Powerpoint presentation posted: http://www.iassistdata.org/conferences/2004/presentations/F1_hu.ppt

⁷ Many of these sites and those following are linked on the University of California, Irvine's Center for the Study of Democracy "Asia Data Archive" web page: <http://www.democ.uci.edu/resources/asia.php>.

Cultural barriers

In Chinese there is a word that perhaps summarizes the ultimate challenge to data sharing in China and other Confucian-based cultures in Asia: Guanxi. According to this perspective, unless one can establish guanxi, nothing can be done. There has to be some personal connection – or interpersonal trust -- before a data owner will share his or her data with you.

As new emerging economies join the world market, there is an increasing need for data. IGOs and increasingly, NGOs, pour money in to conduct surveys. Yet the culture of NGO work is such that all the donors require is a printed, preferably glossy report. There is no demand as of yet for the raw data, so data files get stored haphazardly and are not systematically archived.

Financial barriers.

Lack of incentives to share or preserve

While national social science academies have been established throughout Asia, there is little support for researchers seeking raw data for their analysis. In Vietnam, for example, researchers who attend conferences abroad are reimbursed for book purchases, but given the focus on researching local conditions, there is little if any interest in raw data generated abroad, which in any case may be quite costly to acquire. Three Asian countries – China, Japan and South Korea -- do have national memberships in ICPSR, and there are individual institutions in Asia which are ICPSR members (including University of Hong Kong and Academia Sinica in Taiwan) but no institution in Southeast Asia, for example, has joined.⁸

Political barriers

One cannot minimize political barriers to data collection, preservation and dissemination. Politically sensitive questions in cross-national surveys are omitted in particular countries. Or surveys that have results that are uncomplimentary to the country's government may be suppressed. The POP center's experience here at the University of Hong Kong is illustrative but still mild compared with what could occur across the border or elsewhere in the region. Our host Robert Chung's own past experience with political interference is well-known⁹, but one can imagine a scenario where data about universal suffrage protests (such as the one that occurred earlier this month) might cause the central government to

⁸ ICPSR Member Institutions: <http://www.icpsr.umich.edu/membership/ors.html>

⁹ See documents relating to the University of Hong Kong's investigation panel: http://jmsc.hku.hk/HKUPanel/transcripts/report/PollHearing_Report_01.htm

become queasy and want not to have it disseminated to researchers within China, given the apparent news blackout¹⁰ on the mainland about the recent huge protests.

Yet technological innovations make data sharing much more likely, and the barriers I've enunciated less relevant in some cases, or easier to overcome.

One remaining major barrier is implementing multilingual access to those whose first language is not English. While those trained in English may not see the domination of English in Internet-communication as a barrier, there is a danger that such reliance on English makes the web much less multicultural, hastening the erasure of local distinctiveness. Nishigaki Toru has argued this much more eloquently so I quote him when he explains the limits of English being the sole way to communicate in cyberspace:

[I]t will be a rather unhappy option as far as multiculturalism is concerned. For many people the need to communicate via English will essentially cut them off from their own cultures and traditions, leaving them rootless and adrift in cyberspace. Correspondingly, the threat this poses to native cultures might anger a number of other people enough to inspire a dangerous reaction, a turn toward more virulent forms of racism, nationalism or xenophobia.

We need to take time to ensure that the Internet provides a space from where people from different cultures and value systems, including those from the United States, can freely and equally exchange ideas with one another into the 21st Century: the development of a multilingual environment is therefore indispensable...¹¹

For data providers such as pollsters then, there are many challenges, not just the need to provide adequate translations of questionnaires and variable names and responses.

Metadata transformation

Currently social science data archives and their archivists and librarians have been at the forefront of making data more accessible. Among them is the development of a new standard for data documentation, DDI.

Data providers such as pollsters may not be aware of this, but this new standard has been revolutionizing the way one can do cross comparisons across time and across polls. This new standard in essence makes every line in a codebook searchable because DDI – the Data Documentation Initiative – gives the rules on how to encode metadata in XML.¹²

¹⁰ Alice Yan, Kevin Huang and He Huifeng, "News of Protest Fails to Surface on Mainland," *South China Morning Post*, December 6, 2005, A3.

¹¹ Nishigaki Toru, "A Multilingual Information Environment for a Multicultural World," *Social Science Information Japan* 10 (August 1997), 8.

¹² See webpage: Data Documentation Initiative: A Project of the Social Science Community: <http://www.icpsr.umich.edu/DDI/>

Data encoded according to DDI facilitates searchability – thus enabling comparability. Technological advances – such as distributed computing and cheap digitized storage (servers) now allow researchers doing secondary analysis of previously collected surveys to sit at a “scholar’s workstation” and with several clicks grab variables across surveys.

One project that uses open source technology with the ability to serve up datasets for public access is the VDC – Virtual Data Center¹³ at Harvard. It remains to be seen how many universities will implement the software (UCI will be testing it).

Data Ferret¹⁴ has been used by the U.S. Census Bureau to serve up survey data across data producers. Users need to download an applet or application to make it work.

Nesstar¹⁵ is a commercial project that emerged out of the social science data community in Europe. ICPSR has recently begun offering some of its datasets via Nesstar (Scroll down to List All Nesstar studies link on this page: <http://www.icpsr.umich.edu/access/sda.html>)

SDA, or Survey Documentation & Analysis, a tool developed at UCI Berkeley, does not require DDI compatibility but it also serves up datasets for online analysis one by one. ICPSR has served up selected datasets with SDA online analysis functionality, including some polling data (<http://www.icpsr.umich.edu/access/sda.html>) UCI has been testing it with test data from the POP center (<http://data.lib.uci.edu/hkpolls>).

Conclusion

More must be done to encourage the development of data archives within Asia. Many of the existing archives or project-base datasets are archived in the West, especially the United States. For the immediate future, many projects will continue to be funded with external funds, either from IGOs or NGOs. Donor agencies can require all future polls to be deposited in a publicly accessible archive, either in-country or abroad. But the donors need also to commit funds to support data archiving activities, which while not as “sexy” as the production of glossy reports, will do more to advance the research value of such public opinion data than anything else. It also will help preserve the historical legacy of public opinion in a country, which would otherwise be lost forever.

¹³ “Virtual Data Center: An Operational, Open-source, Digital Library to Enable the Sharing of Quantitative Research Data” -- <http://thedata.org/>

¹⁴ Data Ferret: For the DataWeb: <http://dataferrett.census.gov/>

¹⁵ Nesstar: <http://www.nesstar.org/>