Social Science Data Sharing & Management in Asia: Overview and Future Prospects

By

Daniel C. Tsang

Data Librarian
University of California, Irvine

Powerpoint slides prepared for presentation at
Pacific Rim Digital Library Alliance meeting,
Los Angeles, 10 November 2011
Copyright © 2011

ICPSR Data at UCI circa late 1980s…

Why Share Data?

- Build scientific knowledge
- Create community of data users
- Researchers get access remotely
- Citations may increase
- For publicly funded research, the right thing to do
- Mandated by funding agencies
- Expand storage locations

Social Sciences Data-based Research

- Faculty know the importance of data sharing
- Also, the importance of secondary data analysis
- Key individual faculty instrumental in making data available

Data Sharing in Hong Kong

In Hong Kong in June 2010, Hong Kong University organized a forum on Research Data Sharing

to discuss what Hong Kong needed to do to preserve and archive government-funded research data.

Among the speakers, HKU Prof. John Bacon-Shone gave a presentation on making research data in Hong Kong publicly accessible.

Source:

http://www3.hku.hk/rss/index.php/links-resources/spotlight/forum-on-research-data-sharing

Danger of Losing Datasets

- Earlier this year [2011], he advocated the archiving of research data and warned that already important data have been lost:
- 1976 ByCensus data lost forever!
- 2001 Millennium Sport Survey disappeared

Source: Bacon-Shone, John. “Archiving Research Data in Universities” (Powerpoint slides): http://www.lawtech.hk/wp-content/uploads/2011/01/Session-II-1415-1440-John-Bacon-Shone.pdf

Universities Service Centre of China Studies, CUHK

About the USCCS

- With a view to promoting quantitative research on contemporary China, the Universities Service Centre of China Studies decided in 1995 to start collecting machine-readable data. There are a number of hurdles to overcome in the process. First of all, it is pioneering to make electronic data in China available to public use. Secondly, it is almost impossible to convince people there of the idea of data archiving and data-sharing for academic purpose, as information is sometimes regarded as intelligence, power or commercial goods. Finally, we succeeded in establishing several frameworks of academic cooperation in data development; but when the data arrived, we had to spend considerable man-hours to perform quality control. Frustrations often ensued as errors could not be corrected for one reason or another. We only hope that as quantitative research skills and professional attitudes improve, the quality of data will also be enhanced. The USC-Databank for China Studies can also make a humble contribution towards the rise in standard.

Faculty Involvement Crucial

We are committed to collaborative efforts to enable data-sharing for academic study. We wish to thank Prof. Tom Lyons who was the first scholar to have gracefully deposited the datasets on Economic Geography of Fujian of the"Cornell East Asia Series" in our holdings. We hereby appeal to individuals and institutions to follow his step. Only with your generous support, we may be able to better serve the global academic community with more quality datasets at lower costs. - From Universities Service Centre of China Studies web site: http://www.usc.cuhk.edu.hk/Eng/AboutDCS.aspx

We are pleased to announce that the English version of "China Survey Data Network" (CSDN) is now ready for release. As a joint effort by the China Data Center of the University of Michigan and the China Center for Economic Research of Peking University with an initial funding support from the Ford Foundation Beijing office, CSDN is a web-based data network for the archiving of microdata on China. Its primary mission is to promote the sharing of microdata on China among researchers within and outside of China.

Date: 2009-05-06

We are pleased to announce that the English version of ¡°China Survey Data Network¡± (CSDN) is now ready for release. As a joint effort by the China Data Center of the University of Michigan and the China Center for Economic Research of Peking University with an initial funding support from the Ford Foundation Beijing office, CSDN is a web-based data network for the archiving of microdata on China. Its primary mission is to promote the sharing of microdata on China among researchers within and outside of China.

CSDN has a bilingual website at http://www.chinasurveycenter.org. The main functions on the current web site include Find Datasets, Data Deposit, Data Tutorial, and Online Survey. The web site also includes many links to other online resources for China studies. If you have a microdata set on China that you are willing to share with others, we encourage you to deposit the data through the CSDN website.

Self-registration is now open for all potential contributors or users of microdata on China. We invite you to visit the CSDN web site at http://www.chinasurveycenter.org. Please feel free to share this message with your colleagues who might be interested in it. We look forward to your contributions to CSDN as well as your feedback on how CSDN can be improved. Please feel free to contact us at csdn@ccer.edu.cn or chinasurvey@umich.edu if you have any suggestions.

Please feel free to forward this to your colleagues who might be interested in this.

CSDN Executive Committee:

Albert Park, Department of Economics, University of Oxford
Shuming Bao, China Data Center, University of Michigan
Yang Yao, China Center for Economic Research, Peking University

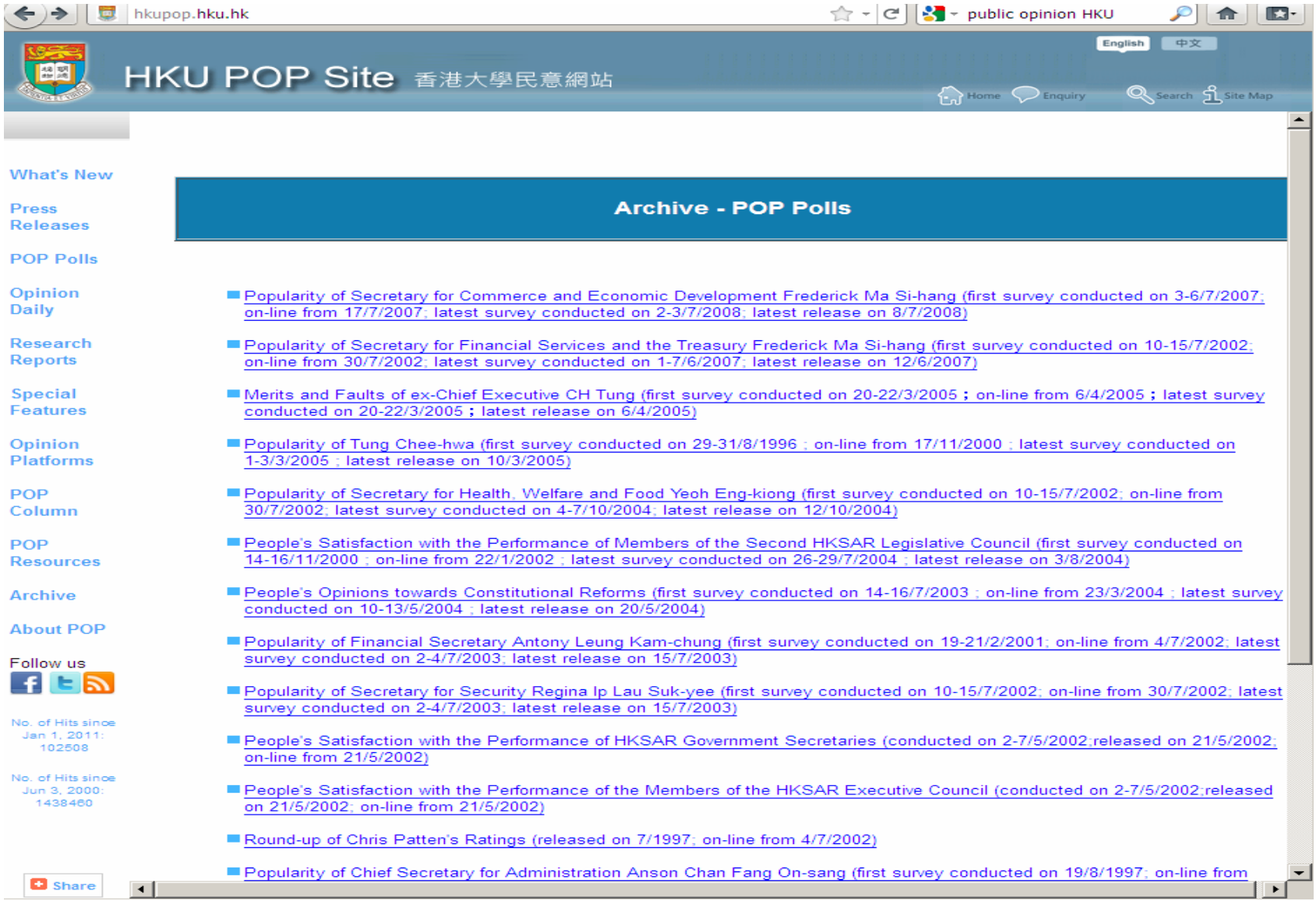
China Survey Data Network
U. of Michigan & Beijing University

Downloadable data include:

- Poverty and Development: CCAP Village Survey (2003)
- Poverty and Development: CCAP Village Survey (2000)
- Jiangsu Rural Household Survey （2002）
- Chinese College Graduates Employment and Skills...
- China Urban Labor Survey (CULS, 2001)
- Private Sector Survey (2006)
- Impacts of Property Rights Reforms: SOEs and TVEs...
- Gansu Survey of Children and Families, first wave

http://www.chinasurveycenter.org/csdn_en/

Public Opinion Programme, HKU



ICPSR | INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH

China Multi-Generational Panel Dataset, Liaoning (CMGPD-LN), 1749-1909

Principal Investigator(s): Lee, James Z.; Campbell, Cameron D.

Summary: The China Multi-Generational Panel Dataset - Liaoning (CMGPD-LN) is drawn from the population registers compiled Agency (neiwufu) in Shengjing, currently the northeast Chinese province of Liaoning, between 1749 and 1909. It provides 1.5 m more than 260,000 residents from 698 communities. The population mainly consists of immigrants from North China who settle early eighteenth century, and their descendants. T... (view details)

Persistent URL: http://dx.doi.org/10.3886/ICPSR27063.v6

Access Notes

- One or more files in this study are restricted; consult the detailed metadata page to learn more. You can apply for acces Portal. A MyData account is required to apply for access.
- This study was originally processed, archived, and disseminated by DSDR, a project funded by the Eunice Kennedy Shrive (NICHD).

What Can I Do With This Collection?

- Explore
  - View study description
  - Browse documentation files (login not required)
  - Variables:
    - List all variables in this study
    - Search the variables in this study

      Search for Variables
  - View related literature (~48)
- Download

China Multi-Generational Panel Dataset

Home About Us Data Publications Training Documentation Contact Us

Welcome

The China Multi-Generational Panel Dataset - Liaoning (CMGPD-LN) has 1.5 million triennial observations of more than 260,000 residents from approximately 698 communities in the northeast Chinese province of Liaoning between 1749 and 1909. The data provide socioeconomic, demographic, and other characteristics for individuals, households, and communities, and record demographic outcomes such as marriage, reproduction and death. (more)

Announcements

- CMGPD-LN analytical and kinship datasets available
- Video Presentation from "Public Health in Asia" Event
- New Article
- CMGPD Training Workshop
- Restricted data available

About the Researchers

Cameron Campbell is Professor of Sociology at UCLA. He received his BS from Caltech, his MA and PhD from the Un postdoc at the University of Michigan Population Studies Center. His research focuses on the relationships between demographic behavior. He has published extensively on family and population in eighteenth and nineteenth century *and Fortune in Rural China* with James Lee. Recently he has published papers on ethnic identity and social mobility, participant in the Eurasia project, an international collaboration that compares relationships between economic con behavior for a variety of historical European and Asian communities. He is co-author of a volume from this effort, *Li* examines how household responses to economic stress were reflected in mortality patterns. With James Lee, he is and kinship in northeast China from the seventeenth century to the present.

Shuang Chen is Assistant Professor of History at the University of Iowa. She received her BA and MA in History fro the University of Michigan. She joined the University of Iowa in August 2010, after spending a year as a post-docto project. Professor Chen specializes in the social, economic, and political history of late imperial and modern China. settlement, population behavior, and social stratification. As a member of the Lee-Campbell research group, she ha Campbell on the analysis of historical demographic and socioeconomic data drawn from population and land registe manuscript on interactions between wealth stratification, demographic processes, and institutional context that is b settlement history and the subsequent evolution of inequality in land ownership in Shuangcheng County, Heilongjia century.

James Z. Lee (1952 -) is Dean of the School of Humanities and Social Sciences at the Hong Kong University of Scie Associate at the Inter-University Consortium for Political and Social Research at the University of Michigan, and Cha Peking University. A practitioner of social scientific history, the application of quantitative social science methods o colleagues in the Lee-Campbell Research Group link historical and contemporary archival sources, social surveys, ge oral histories to create large individual level panel data sets extending from late imperial to contemporary China. T despite recent profound political, social, and economic changes, many distinctive institutions and patterns of demo and social mobility persist from China's imperial past.

Professor Lee's published work includes six authored or co-authored books, five co-edited books, and fifty articles fo demography, ethnicity, fiscal and frontier history of late imperial China, as well as on the social organization, and s and contemporary China. He has recently extended his area of research from historical China to the comparative de other East Asian and West European populations in the past and China and Zambia in the present. A John Simon G Science History Association's Allan Sharlin Award for Best Book in Social Science History (2000), and two American S Duncan Award for Distinguished Scholarship in Social Demography from the Population Section (2000) and the Outst and Asian America Section (2005). His book, La population chinoise. *Mythes et réalités* (Presses de l'Université de M Falardeau for best French-language book in the Social Sciences from The Canadian Federation for the Humanities an *Reproduction and Human Agency in Europe and Asia, 1700-1900* was just published by MIT Press in 2010.

Susan Hautaniemi Leonard is an Assistant Research Scientist at the Inter-University Consortium for Political and S Studies Center, University of Michigan. Dr. Leonard's work focuses on the relationship between human populations a

ELSEWHERE in ASIA…

In 2004 I wrote about Vietnam, where I did some research on its data developments as a Fulbright Scholar. The appeared in the Spring 2004 issue of *IASSIST Quarterly*, a peer-reviewed, open-source online publication from IASSIST the longstanding, international community of data archivists, data librarians and data producers, that has encouraged cross fertilization across the world, and where I was active as an administrative committee member for eight years and also as an advocate for data sharing and involvement from Asia.

Its forthcoming annual conference will be in Washington D.C. June 4-8, 2012.

IASSIST can be reached at http://www.iassistdata.org

Reflections on a Quest for Social Sci Data in Vietnam

by Daniel C. Tsang [1] *

ing country likely to
ld Trade Organization in
Vietnam is, perhaps not
wash in social science
vernment mandates,
projects, and new research
contribute to a renewed
untry seeking very hard
obal mainstream. Indeed,
in my Fulbright research
project statement" drafted in 2002:

moves toward entry into the World Trade
and its economic system becomes more
science research data can be expected to play
y important role for scholars interested in
and analyzing Vietnamese society, as well as
ers within Vietnam and those in the United
might add here, elsewhere.]. Scholars and
will also be interested in using such data for
lysis.

I mentioned that V
social science stud
of that is the prolif
country's General
(GSO), which pub
year thick compen
statistics, gathered
and administrative

Annually, it publis
(Vietnamese/English) *Statistical Yea*
in a CD-ROM edition as well. One
able to pick up was *Major Socio-eco*
Obtained from Ten Large Scale Surv
2000. Hanoi: Statistical Publishing
heavy volume; it ran over 2,000 pag
from not only the ten different surve
socio-economic indicators of 61 pro
chapter headings for the ten surveys
some idea of the proliferation of sur
years in Vietnam:

Need for in-country Data Archives

- …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.

Source: Tsang, Daniel C., "Reflections on a Quest for Social Science Data in Vietnam," *IASSIST Quarterly* 28/1 (Spring 2004), 20 http://www.iassistdata.org/downloads/iqvol281tsang.pdf

Rapid Change

All that is rapidly changing, in just less than a decade.

In Vietnam, for example, thanks to the concerted efforts of many international data folks, the General Statistics Office has placed an online Data Warehouse for two of its largest time series data surveys and is about to populate a web-based Data Archive for publicly available datasets that will be freely downloadable from its site.

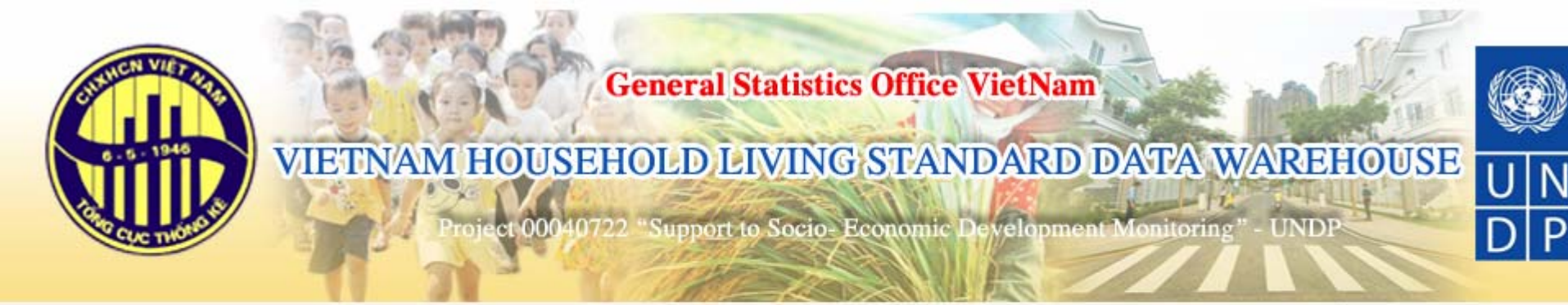
International Data Sharing Efforts

Two prime movers and shakers:

The International Household Survey Network, for providing the tools and guidelines used to establish National Data Archives.

The Accelerated Data Program from the World Bank and PARIS21, for providing technical and financial support for its establishment.

I provide several examples from Vietnam on the next slides.



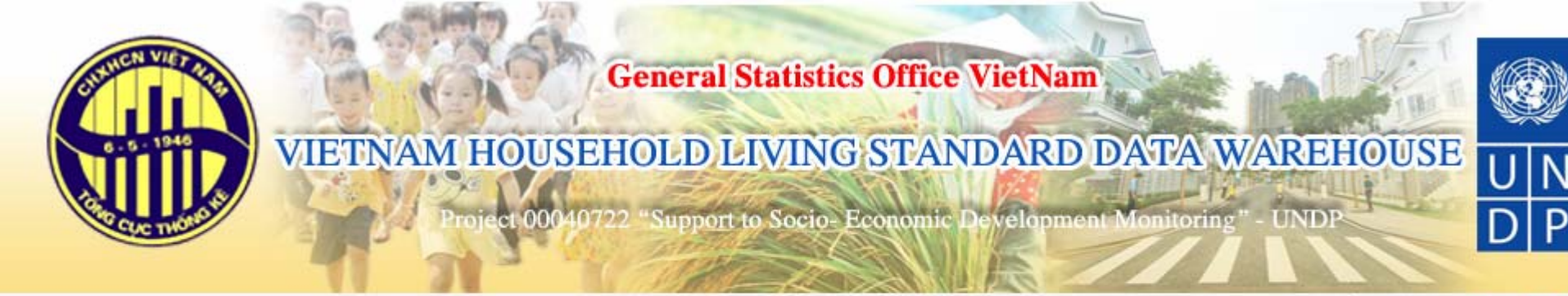
INTRODUCTION
Overview
Survey plan
Questionnaire

AGGREGATED STATISTICS
Demography
Education
Labor and employment
Health
Income
Accommodation and land
Fixed assets and durable appliances
Expenditure
Poverty

META DATA

INTRODUCTION/ Overview

PROJECT 00040722 "Support to Socio-Economic Development Monitoring" supported by UNDP and DFID, aims to strengthen national monitoring framework through the delivery of revised/updated National Statistical Indicator System necessary for So economic development planning and SEDP/VDG/MDG monitoring; revised/updated National Statistical Action Plan; improved quality and harmonized data collection; and improved data reporting, communication, use and storage. One of the most impo tasks of the Project is to support GSO to develop a place to store data with mining tools for different data users to improve data acc In order to achieve this objective, GSO needs to develop statistical data warehouses to facilitate more effective management using. However, since development of statistical data warehouses is new to GSO, many activities were carried out such as conduc study tours to some countries, and evaluating data sources. Minh Viet Company (an IT firm) was recruited to carry out an i problem definition and requirements analysis. On the other hands, GSO conducted some study tours to Korea, Canada, etc to l experience and practices. The study tour to Statistics Canada was headed by Mr. Nguyen Duc Hoa, Vice Minister of Planning Investment cum Director General of GSO. The knowledge and experience obtained from these study tours have facilitated GS develop a statistical data warehouse for Viet Nam Household Living Standard Surveys (VHLSSs) 2004, 2006 and 2008 with the a participation of Center for Statistical Information Technology and Socio-Environmental Statistics Department.



INTRODUCTION
Overview
Survey plan
Questionnaire

AGGREGATED STATISTICS
Demography
Education
Labor and employment
Health
Income
Accommodation and land
Fixed assets and durable appliances
Expenditure
Poverty

META DATA

AGGREGATED STATISTICS >> Health

- Average healthcare expenditure per person having treatment in the past 12 months
- Average of inpatient healthcare expenditure per person having treatment in the past 12 months
- Average of outpatient healthcare expenditure per person having treatment in the past 12 months
- Expenditure except treatment in the last 12 months
- Expenditure for treatment per capita in the last 12 months
- Monthly healthcare expenditure per capita in the last 12 months
- Number of days off due to the illness or injuries per person being ill or injured
- Payment of medical instruments
- Payment of medicines for self-treatment or reservation
- Payment of volunteer health insurance
- Percentage of inpatient treatmentn having health insurance or free health care certificate
- Percentage of inpatient treatmentn in the past 12 months
- Percentage of outpatient treatment having health insurance or free health care certificate
- Percentage of outpatient treatment in the past 12 months
- Percentage of people having treatment in the last 12 months
- Percentage of persons suffered from illness or injuries in the past 12 months
- Percentage of persons suffered from illness or injuries in the past 4 weeks
- Percentage of persons suffered from illness or injuries who have to stay in bed and need some one to take care

INTRODUCE

Overview

Survey plan

Questionnaire Survey

GENERAL TARGETS

1. General population

2. Population 15 years and over

3. Population in working age

4. Population 15-34 years

Meta Data

Concepts, definitions, calculation methods

List of classification

List of tables

Classified variables

Data on labour and employment plays an increasingly important role in evaluating the labour market in particular and the national economy in general. Since 2007, the General Statistics Office of Vietnam (GSO) has conducted the Labor Force Surveys annually for this purpose.

In order to systematize the information and results of the Labor Force Surveys conducted in 2007, 2008, 2009 and 2010, a data warehouse of the above-mentioned surveys has been built to meet requirements of data users in the country and abroad.

We sincerely thank United Nations Development Programs (UNDP) for providing technical and financial supports to build the data warehouse on labour and employment. We also highly appreciate the dedication and hard working of the staffs of the Department for Population and Labour Statistics and the Center for Statistical Information Technology N°I of the General Statistics Office to contribute to the completion of the data warehouse.

It should be noted that population size of the whole country and of provinces/cities during the period of 1999-2009 have been adjusted according to the completed result of the 2009 Population and Housing Census. For this reason, weights of the Labour Force Surveys have been recalculated. Therefore indicators in this data warehouse might be slightly different from those released before.

Despite efforts, mistakes are inavoidable in this data warehouse. We hope to receive feedbacks from data users so that the data warehouse is more improved in terms of quality.

General Statistics Office

July, 2011

Home
Policies and procedures
Acknowledgements
Data catalog
Our mission
Our activities

Home

Survey and census microdata are invaluable resources for statisticians, researchers and analysts. They constitute important and irreplaceable assets which need to be managed in a way that encourages their use and re-use, while protecting the privacy of respondents.

The National Data Archive has been established to:

- Promote best practice and international standards for the documentation of microdata amongst data producers in the country.
- Provide equitable access to microdata in the interest of all citizens, by protecting confidentiality and following international recommendations and good practice.
- Promote the effective use of existing survey and census data for statistical and research purposes, thereby encouraging a diverse range of analytical work through secondary research.
- Ensure the long term preservation of microdata and the related metadata, and their continued viability and usability in the future.

The National Data Archive pursues these objectives within the framework of the national legislation and of the United Nations Fundamental Principles of Statistics. Where microdata cannot be released due to confidentiality or other reasons, the National Data Archive provides the public with detailed metadata and other publicly available materials.

Home
Policies and procedures
Acknowledgements
Data catalog
Our mission
Our activities

Policies and procedures

Micro-datasets are categorized into three groups, according to the sensitivity of their content and their inherent disclosure risks:

- Public use files
- Licensed files
- Files accessible on-site (data enclave)

Public use files

Public use files are microdata files characterized by a low disclosure risk, or by the fact that respondents gave their formal consent to the data dissemination. They are made available on-line to all interested users, for research and statistical purposes only. Users are asked to complete an on-line form and agree to abide by our terms and conditions before being provided with the data.

Licensed files

This approach involves a signed agreement between the National Data Archive and external trusted users, to permit them to access semi-anonymized data files. Licensing agreements are only entered into with *bona fide* users when there is an appropriate sponsoring institution or referee. Users must be able to:

- Demonstrate a need to access the data in order to fulfill a stated statistical or research purpose; and
- Comply with the conditions set forth in a formal Microdata Access Agreement.

Those interested in accessing the data under license should fill our *Licensed Dataset Request Form*. Requests are evaluated by a our Microdata Release Panel. When the decision is made to grant access to the data, users will receive them through secure ftp server.

Files accessible on-site (data enclave)

Where the data is particularly sensitive, access is only provided on-site in our data enclave under strict conditions, and only for research purposes. The data enclave is located in [location: agency, city, country]. The computers within the enclave are not linked to the outside world: researchers do not have email or internet

Social Science Data

Resources for locating and analyzing social science data across subfields and geographic regions

Asia

For Vietnam, click here.

- AsiaBarometer
  Data on public opinion in Asia, covering East, Southeast, South and Central Asia. Three waves of the AsiaBarometer have be
  Focus is on daily lives of ordinary people.

- Australian Social Science Data Archive
  Based at The Australian National University (ANU),this data archive was established in 1981. Datasets available upon request.
  completion of a form is required before downloading of data. The data catalog is browsable.

Asia: Vietnam

- Demographic and Health Surveys: Vietnam ☆☆☆☆☆
  Access to survey documentation and data (after registration & approval) from 1997 and 2002 surveys in Vietnam. Site also lin
  surveys elsewhere.
- Vietnam Life History Survey, 1991 ☆☆☆☆☆
- Vietnam Longitudinal Survey, 1995-1998 ☆☆☆☆☆
- Vietnam Living Standards Survey (VLSS), 1997-98 ☆☆☆☆☆
- Vietnam Development Information Center ☆☆☆☆☆
  [Gateway to development-related research in Vietnam]
- Vietnam Trade Database ☆☆☆☆☆
- International Data Resource Center: Vietnam data ☆☆☆☆☆
  Clearinghouse to international data. This link indexes data from Vietnam gathered in various surveys.
- World Values Survey 2001: Vietnam ☆☆☆☆☆
  Papers related to the Vietnam component of the World Values Survey 2001. The dataset itself is available from ICPSR.
- General Statistics Office of Vietnam ☆☆☆☆☆
  Statistical data and publications for the entire country
- Statistical Office in Ho Chi Minh City ☆☆☆☆☆
  A branch of the General Statistics Office, most of the site is in Vietnamese with some data tables in English and Vietnamese
- Integrated Public Use Microdata Series — International ☆☆☆☆☆
  Includes sample census from 1989 and 1999 Vietnam census. Free access but must apply.
- Vietnam Northern Uplands, Qualitative Social Assessment of Income Diversification, 2002 ☆☆☆☆☆

SDA [Use classic interface] Selected Study: 2000-2001 WORLD VALUES SURVEY: VIETNAM

Analysis | Download | Codebook | Search | Getting Started

Variable Selection: Help

Selected: View

Copy to: Row | Col | Ctrl | Filter

Mode: Append Replace

Search: Go

- 2000-2001 WORLD VALUES SURVEY: VIETNAM
  - CASEID - interview number
  - V1 - wave
  - V2 - nation
  - V4 - family important
  - V5 - friends important
  - V6 - leisure important
  - V7 - politics important
  - V8 - work important
  - V9 - religion important
  - V10 - Service important
  - V11 - is r happy
  - V12 - state of health
  - V13 - respect parents
  - V14 - parents duty
  - V15 - child qual:indep
  - V16 - child qual:hard work
  - V17 - child qual:respons
  - V18 - child qual:imagin
  - V19 - child qual:tolerance
  - V20 - child qual:thrift
  - V21 - child qual:determin
  - V22 - child qual:relig
  - V23 - child qual:unself
  - V24 - child qual:obedience
  - V25 - people trusted
  - V26 - People take advantage

SDA Frequencies/Crosstabulation Program

Help: General / Recoding Variables

REQUIRED Variable names to specify

Row:

OPTIONAL Variable names to specify

Column:

Control:

Selection Filter(s): *Example: age(18-50)*

Weight: No Weight

TABLE OPTIONS

Percentaging:
- ☑ Column ☐ Row ☐ Total
- ☐ Confidence intervals Level: 95 percent
- ☐ Standard error of each percent

N of cases to display:
- ☐ Unweighted ☑ Weighted

- ☐ Summary statistics
- ☐ Question text ☐ Suppress table
- ☑ Color coding ☐ Show Z-statistic
- ☐ Include missing-data values

CHART OPTIONS

Type of chart: Stacked Bar Chart

Bar chart options:
- Orientation: ◉ Vertical ○ Horizor
- Visual Effects: ◉ 2-D ○ 3-D

Show percents: ☐ Yes

Palette: ◉ Color ○ Grayscale

Size - width: 600 height: 400

Title:

Run the Table | Clear Fields

Change number of decimal places to display

NSF Data Management Plan

As of January 18, 2011, all National Science Foundation (NSF) grant applications require a supplementary document of no more than two pages labeled "Data Management Plan". This supplement should describe how the proposal will conform to NSF policy on the dissemination and sharing of research results, and may include:

- types of data (including samples, physical collections, software);
- metadata standards to be used;
- policies for access and sharing (including provisions for privacy/intellectual property);
- policies and provisions for re-use; and
- plans for archiving and preservation of access.

Recommended documents from the UCI Office of Research that support submission of the NSF Data Management Plan include:

1. Suggested Elements to Cover in a Data Management Plan
2. Data Management Plan: Template with Suggested Content
3. NSF Data Management Plan

See Grant Proposal Guide (GPG) Chapter II.C.2.j for full implementation information.

Send questions regarding the NSF Data Management Plan to: DATAMP@uci.edu.

The UCI Libraries stand ready to assist grantees in identifying the options for sharing research data in an institutional or discipline-specific repository. Contact your subject librarian or the UCI Data Librarian (dtsang@uci.edu).

Policies & Guidelines

Investigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants. Grantees are expected to encourage and facilitate such sharing in the NSF Award & Administration Guide (AAG) Chapter VI.D.4.

Requirements by Various NSF Units

Links to data management requirements and plans relevant to specific Directorates, Offices, Divisions, Programs, or other NSF units, are provided below. If guidance specific to the program is not provided, then the requirements established in Grant Proposal Guide, Chapter II.C.2.j apply. Particulars for the plan will depend on what NSF refers to as the "community of interest" -- essentially the domain or discipline to which the project proposal is relevant.

Please note that if a specific program solicitation provides guidance on preparation of data management plans

University of California Curation Center

Merritt
EZID
Web Archiving Service

Data Management Plan

- Types of Data
- File Formats
- Organizing Data Files
- Persistent Identifiers for Data
- Sharing and Archiving Data
- Data Documentation and Metadata
- Citing Data
- Copyright and Privacy/Confidentiality
- Security/Storage/Backups
- Funding Guidelines
- Credits
- UC3 contacts

UC3

Manage Your Data

Managing your data before you begin your research and throughout the research life cycle is essential to ensure usability, preservation and access. Federal agencies and other funders now require that grant awardees include a data management plan with their grant proposals. We can help you meet these new requirements.

UC3 is ready to consult with UC faculty members and researchers as you develop data management plans. Please contact UC3 directly, or contact your UC campus representative for more information.

Data Planning Checklist:

A. Your data management plan
B. Funding agency requirements
C. Creating your data
    1. What types of data will be produced for your project?
    2. What identifiers will you use for your data?
    3. How will you document your data?
    4. How much data will the project produce?
    5. How often will the data change or be updated, and will versions need to be tracked?
D. Organizing your data
    1. What file formats will be produced for your project and what kinds of data management risks do they present?
    2. How will you organize your files into directories and what naming conventions will you apply to both?
E. Managing your data
    1. Who is responsible for managing and controlling the data?
    2. For what or whom are the data intended?
    3. How long must the data be retained?
    4. How secure are the data? Do you have a procedure for backing up the data?
F. Sharing your data
    1. Does project funding require your data to be shared or publicly accessible?

DMPTool

Guidance and Resources for your Data Management Plan

The DMP Tool allows you to: 1 2 3 4

Get Started!

Data Management Plan: Sample Plan Created at the DataONE Best Practices Workshop - Santa Fe NM 7/201

Atmospheric CO2 Concentrations, Mauna Loa Observatory, Hawaii, 2011-2013

1. Types of data produced

See a plan created with the DMP Tool

Recent DMP News

DMPs for Gordon and Betty Moore Foundation grants

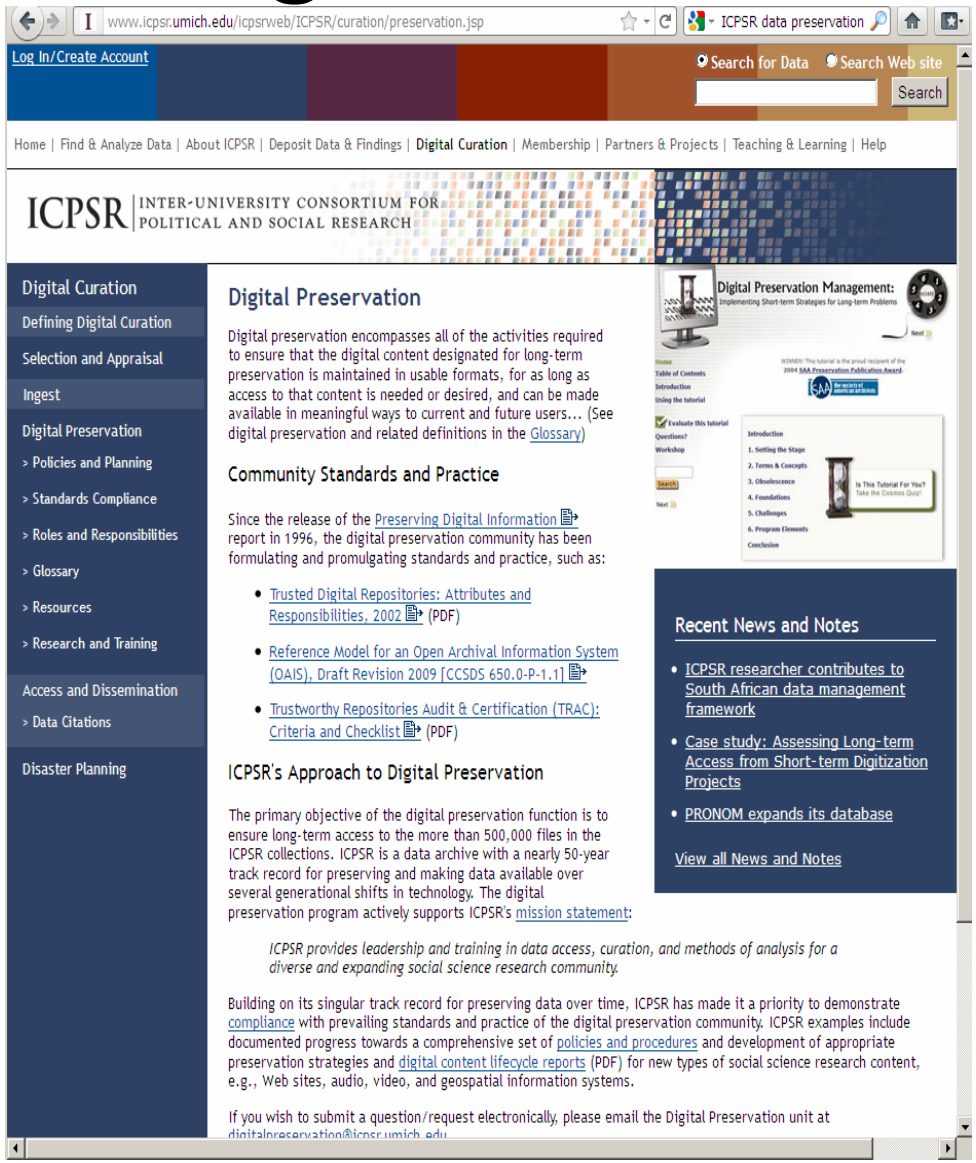
NEH Office of Digital Humanities DMP requirements adde

DMPTool workshop at the DLF Fall Forum

More news >

DMPTOOL is a service of the University of California Curation Center of the California Digital Library
Copyright © 2010-2011 The Regents of the University of California
Privacy Policy | Terms of Use | Photo Credits

ICPSR Digital Preservation



ICPSR | INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH

Digital Curation
Defining Digital Curation
Selection and Appraisal
Ingest
Digital Preservation
> Policies and Planning
> Standards Compliance
> Roles and Responsibilities
> Glossary
> Resources
> Research and Training
Access and Dissemination
> Data Citations
Disaster Planning

Digital Curation

Defining Digital Curation

Digital curation, which encompasses both data curation and digital preservation activities, is the active management and enhancement of trusted digital resources across the life cycle.

Sound curation practices assure that:

- Digital content is captured for long-term use and its integrity assured
- Researchers can find and use digital content for secondary analysis
- Digital content is available in an appropriate form for the designated community
- Privacy is protected for research subjects represented in data
- Digital content is secured in online, near-line, and offline storage
- Digital content is stored in preservable formats for current and future use

Community Standards and Practice

A de facto standard has emerged in the area of digital curation -- the Reference Model for an Open Archival Information System (OAIS), produced by the NASA Consultative Committee for Space Data Systems. OAIS is

The OAIS Data Life Cycle at ICPSR

Preservation planning at ICPSR follows OAIS standards from ingest of digital content to access by data consumers. ICPSR works with data producers to optimize data and documentation, and with consumers to enhance the utility of the data they need.

Click on the image above to see a larger version of the OAIS reference model.

Digital Curation (ICPSR)

- Digital curation, which encompasses both data curation and digital preservation activities, is the active management and enhancement of trusted digital resources across the life cycle.
- Sound curation practices assure that:
- Digital content is captured for long-term use and its integrity assured
- Researchers can find and use digital content for secondary analysis
- Digital content is available in an appropriate form for the designated community
- Privacy is protected for research subjects represented in data
- Digital content is secured in online, near-line, and offline storage
- Digital content is stored in preservable formats for current and future use

Challenges for Libraries

- How to help in the life cycle of a research dataset
- Find out how research data is managed on campus
- Inventory suitable data repositories
- Digital preservation is not just backing it up
- IR is not Data Repository
- Campus IR may not be best for data deposit
- Deal with data & format migration issues
- Deal with access issues

Toolkits & Resources

- DMP Tool: https://dmp.cdlib.org/
- DMP online (UK): https://dmponline.dcc.ac.uk/
- Data Curation Profiles Toolkit: http://www4.lib.purdue.edu/dcp/download
- SDA Resources: http://sda.berkeley.edu/otherarc.htm

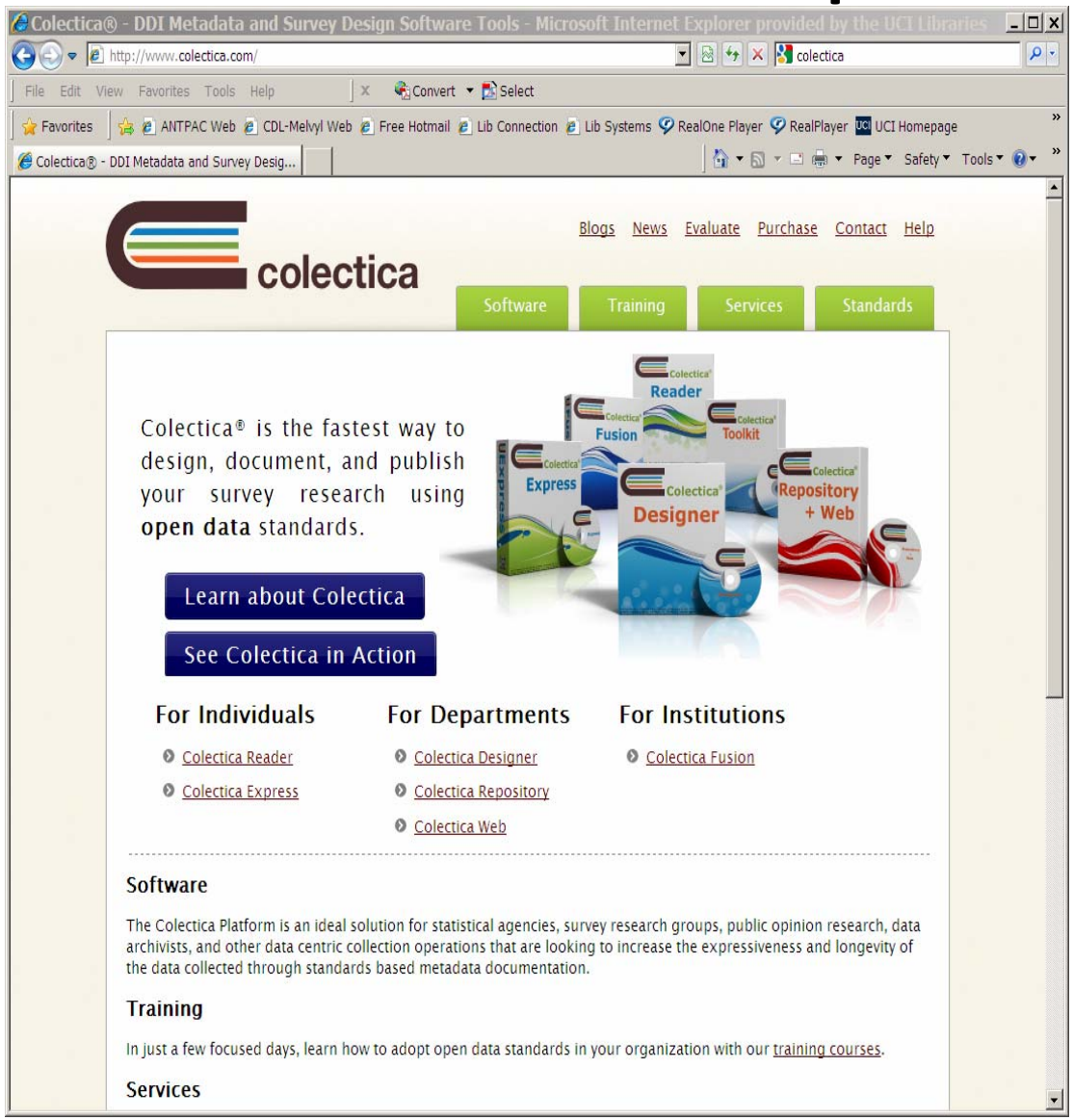
Tooling up and Networking

- Data Curation Profile Workshop, 6 February 2012, UC Irvine (contact dtsang@uci.edu)
- IASSIST: http://iassistdata.org/; upcoming conference, Washington D.C. 4-8 June 2012
- ICPSR Summer Program: http://www.icpsr.umich.edu/icpsrweb/sumprog
- UC3 Webinars: http://www.cdlib.org/services/uc3/uc3webinars.html
- DuraCloud Presentations: https://wiki.duraspace.org/display/DURACLOUD/DuraCloud+Presentations

ICPSR Summer Program in Quantitative Methods of Social Research

Photo credit: ICPSR

Metadata software: A possibility



Conclusion

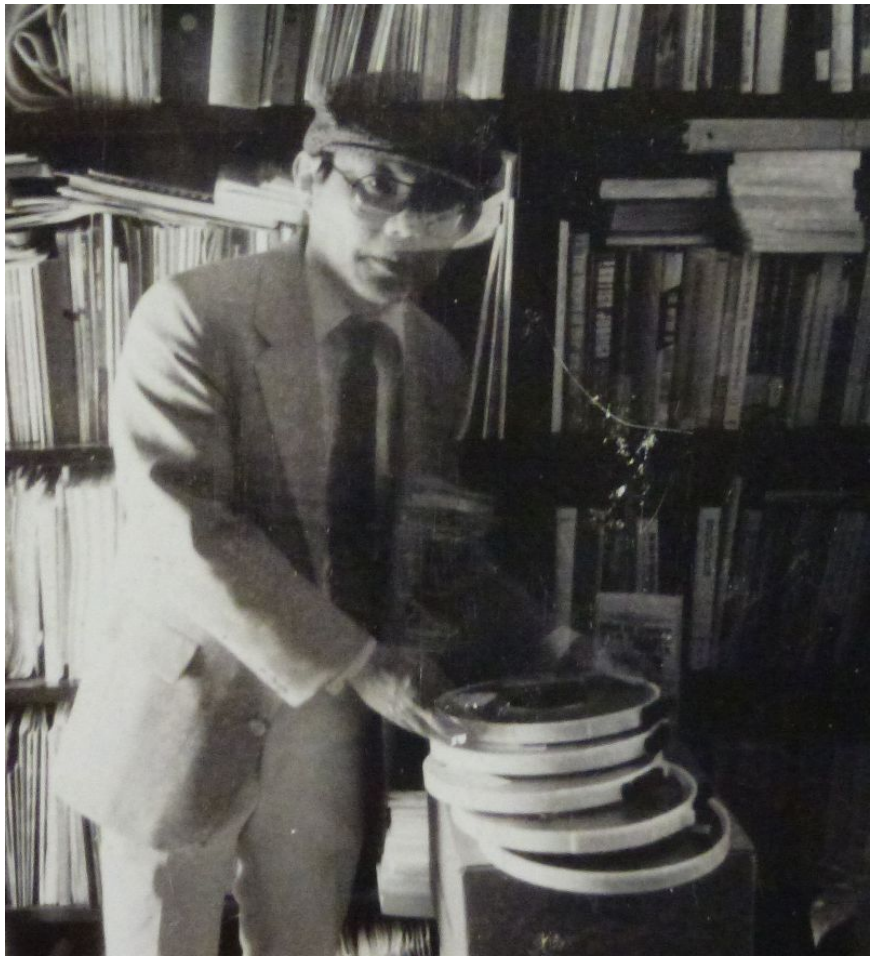
In my 2004 article, " Reflections on a Quest for Social Science Data in Vietnam," I concluded:

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…

We’re Done!



Contact info

Daniel C. Tsang
E-mail: dtsang@uci.edu
1 (949) 824 4978
468 Langson Library
UC Irvine Libraries
PO Box 19557
Irvine CA 92623-9557
USA