Singapore Management University

Institutional Knowledge at Singapore Management University

Library Events SMU Libraries

1-2014

Research Data Management: Challenges and Opportunities

Daniel C. Tsang University of California - Irvine, dtsang@uci.edu

Follow this and additional works at: https://ink.library.smu.edu.sg/lib_events



Part of the Library and Information Science Commons

Citation

Tsang, Daniel C. 2014. "Research Data Management: Challenges and Opportunities." Presentation at Singapore Management University, Li Ka Shing Library, January 7.

This Presentation is brought to you for free and open access by the SMU Libraries at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Library Events by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email cherylds@smu.edu.sg.

Research Data Management

CHALLENGES & OPPORTUNITIES

DANIEL C. TSANG, DISTINGUISHED LIBRARIAN

DATA LIBRARIAN; POLITICS, ECONOMICS &

ASIAN AMERICAN STUDIES BIBLIOGRAPHER

UNIVERSITY OF CALIFORNIA, IRVINE, LIBRARIES

PREPARED FOR PRESENTATION AT SINGAPORE MANAGEMENT UNIVERSITY, 7 JANUARY 2014

E-MAIL: DTSANG@UCI.EDU

University of California, Irvine

- Ranked no. 1 among U.S. universities under 50 years old by Times Higher Education in 2012 and 2013
- Among the top 5 universities worldwide under 50 years old according to Times Higher Education for the past two years
- There have been three Nobel Laureates at UCI
- In 2013, of the 28,895 total student enrollment, 12,048 (or about 42%) are Asian American or Asian permanent residents of the U.S. Seven undergraduates & six graduate students came from Singapore. Two recent Ph.D dissertations [both in Anthropology] done at UCI focused on 1) sexual citizenship among queer Singaporeans and 2) ethical capitalism in Singapore's participation in the global knowledge economy.

Sources: THE World University Rankings: http://www.timeshighereducation.co.uk/world-university-rankings/

Key Rankings: http://www.uci.edu/distinctions/

Student Data: Enrollment: http://www.oir.uci.edu/enrollment.html

Singapore Dissertations Irvine: http://antpac.lib.uci.edu/search/X?SEARCH=singapore+dissertations+Irvine&SORT=D&searchscope=7

UC Irvine Libraries

- Our electronic and print collection includes approximately:
 - ▶ 3.4 million volumes
 - ▶ 137,000 journals and serials
 - ► 138,000 government documents
 - ▶ 136,000 audio/visual materials, multimedia, maps, photographs, and other graphic materials
 - ▶ The Libraries Special Collections preserves 581 archival collections that include over 1 million original documents and photographs

Our local digital repository, UCIspace @ the Libaries, includes many distinct collections, among them, videos and audio (with transcripts) from the Vietnamese American Oral History Project

Sources: The UCI Libraries at a Glance: http://www.lib.uci.edu/about/facts-and-figures.html

UCIspace: http://ucispace.lib.uci.edu/community-list





VAOHP Home

Search VAOHP

Go

Advanced Search

Browse VAOHP

Narrator

Interviewer

Subject

Decade of Birth

Source of Interviews

About VAOHP

About This Project Contact Us

Other Links

VAOHP Blog

UCI Libraries Southeast Asian Archive UCI Dept of Asian American Studies





VIETNAMESE AMERICAN LIFE STORIES OF VIETNAMESE AMERICANS ORAL HISTORY PROJECT IN SOUTHERN CALIFORNIA



▶ This website is a work in progress. Please check back for updates and new oral histories. ◀

University of California Open Access Policy

- ► The Academic Senate of the University of California passed an Open Access Policy on July 24, 2013, ensuring that future research articles authored by faculty at all 10 campuses of UC will be made available to the public at no charge.
- ▶ The policy covers more than 8,000 UC faculty and as many as 40,000 publications a year. By granting a license to the University of California prior to any contractual arrangement with publishers, faculty members can now make their research widely and publicly available, re-use it for various purposes, or modify it for future research publications.
- Faculty on three campuses (UCLA, UCI and UCSF) began depositing articles in eScholarship on November 1, 2013.

DMPTool

https://dmp.cdlib.org/



Contact Us | Get Started | Login

More news >

Stakeholders

An OCLC Research report, "Starting the Conversation: University-wide Research Data Management Policy" (December 2013) lists the following stakeholders for "starting the conversation" about research data management policy on academic campuses:

- ▶ The University
- ▶ The Office of Research
- ► The Research Compliance Office
- ► The Information Technology Department
- ► The Researchers
- ▶ The Academic Units
- ► The Library

Source: http://oclc.org/research/publications/library/2013/2013-08r.html

Elements of the Conversation

- Who owns the data?
- What Requirements are Imposed By Others?
- Which Data Should Be Retained?
- For How Long Should Data Be Maintained?
- How Should Digital Data Be Preserved?
- Are there Ethical Considerations?
- How are Data Accessed?
- ▶ How Open Should the Data Be?
- How Will Costs Be Managed?
- What are the Alternatives to Local Data Management?

Source: http://oclc.org/research/publications/library/2013/2013-08r.html

UC Irvine Libraries Initiatives

- ► The UCI Libraries are partners with the Office of Information Technology (OIT) and the Office of Research in an effort to define the long term direction and priorities for research computing and electronic research services on this campus.
- A committee made up of UCI faculty, and staff from the Libraries and OIT, has made a set of recommendations to campus administration based on an online survey and focus groups with faculty. Among the proposals made are the need to develop a much faster network for the movement of research data across campus and externally; more support staff to enhance services offered, such as management, preservation, and organization of research project data; and development of a research data storage system for long term, secure storage of both raw and processed data sets.

Faculty Assessment of the State of Research Computing (FASRC) at University of California, Irvine (2013)

"Most critical research computing need"

- Our assessment is that long-term research data storage, and associated data management, is the single most critical research computing need not being met on campus.
- ► The FASRC committee believes that a well-run data storage service would allow many faculty groups to coordinate data storage using a centralized system, foster research collaboration, and provide access to archived research data.
- Faculty expressed a need for having a secure place to archive their data, if not centrally, elsewhere on or off campus.
- As a major component of the University's scholarly product, research data must not only be stored securely but preserved and curated in trusted repositories so that the data remain accessible to the research community after a project is completed. Such accessibility enables secondary analysis of research data originally collected by University faculty and researchers

Source: http://sites.uci.edu/fasrc/files/2012/11/Faculty-Assessment-of-the-State-of-Research-Computing-Report.pdf

Finding faculty for potential partnerships

NSF awards to your institution

http://www.nsf.gov/awardsearch/

NIH funded grants to your institution

http://report.nih.gov/award/index.cfm

- Data Management Plans
- Past faculty contributing content to institutional repositories
- Google Scholar
- Data Citation Index
- ORCID

Talking Points with Faculty

UCI Libraries draft in progress

Questions to ask Faculty

- Can you tell me a bit about your research and what sort of data is involved?
- Are you collecting your own or re-using existing data?
- Where is your data currently stored?
- What software and tools do you use to manage or analyze your data?
- ▶ Do you currently share your data? Would you like to share it in the future?
- Do you link your datasets to associated research publications?
- How is your research funded and does the funding agency require data sharing or preservation?
- ▶ Have you completed a Data Management Plan?

Trusted Repositories: The Data Seal of Approval

Data Seal of Approval

- The Data Seal of Approval is one such assessment initiative. Created by the Data Archiving and Networked Services (DANS) archive in The Netherlands and overseen by an international board, the Data Seal of Approval is meant to demonstrate to researchers that data repositories are taking appropriate measures to ensure the long-term availability and quality of data they hold.
- ▶ The seal sets forth 16 guidelines related to trustworthy data management and stewardship. ICPSR was one of the first six data repositories to earn the Data Seal of Approval in 2011. You can read the ICPSR self-assessment here -- http://assessment.datasealofapproval.org/assessment_28/seal/html/. The other five archives awarded the Data Seal of Approval are the Archaeology Data Service (United Kingdom); the DANS Electronic Archiving System (Netherlands); the Platform for Archiving CINES (France); the Language Archive of the Max Planck Institute for Psycholinguistics (Netherlands); and the UK Data Archive.
- The seal is awarded after an online self-assessment regarding a data repository's adherence to the guidelines. The assessment is then reviewed by the DSA Board before the seal is given.
- In Europe, the Data Seal of Approval serves as a Basic Certification step in an <u>integrated</u> <u>framework for auditing and certifying digital repositories</u>.



Data Seal of Approval 2.0 (2013)

Fundamental to the following guidelines are five criteria, that together determine whether or not the digital research data may be qualified as sustainably archived:

- The research data can be found on the Internet.
- ▶ The research data are accessible, while taking into account relevant legislation with regard to personal information and intellectual property of the data.
- ▶ The research data are available in a usable format.
- The research data are reliable.
- The research data can be referred to.

Source: https://assessment.datasealofapproval.org/media/files/DSA_booklets/DSA_bo

Assessment manual: https://assessment.datasealofapproval.org/guidelines_52/pdf/

Purdue University Research Repository

Start Your Research Project

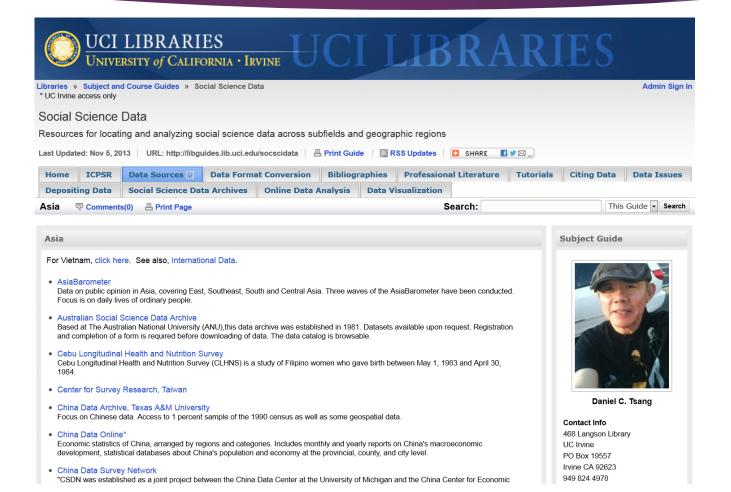
- Create a Data Management Plan: Learn about the detailed requirements for your data management plan (DMP). Funding agency requirements are very specific and our DMP resources can help you to clear up any confusion.
- Upload Research Data to Your ProjectCreate a project to upload and share your data with collaborators using our step-by-step form to guide you through the process. Invite collaborators from other institutions to join your project.
- Publish your DatasetPackage, describe, and publish your dataset with a Datacite DOI. Publishing will ensure your dataset is citable, reusable, and archived for the long-term.

Source: https://purr.purdue.edu/

Developments from UCI Libraries

- Recruitment forthcoming for the head of E-Research & Scholarly Communications
- Potential implementation of DataShare for deposit of and access to research data generated by UCI faculty and researchers
- Collaboration with faculty on archiving research conducted from and around Orange County, especially in the area of biodiversity
- Discussion on using ORCID to identify faculty researchers
- Enhancing subject guides to data sources and data management

Asia Data Sources



Research Data Management



UCI Web Page on DMP

http://nsfdmp.lib.uci.edu/



Challenges to Data Sharing

On the researcher side...

- Data sharing culture varies by discipline & by country
- Sharing data different from sharing published articles
- Not wanting someone to steal your ideas
- Some research, e.g. with personal identifiers cannot be shared
- Repurposing of data may be precluded
- Further consent from respondents necessary
- Privacy of respondents must be protected
- National security may come into play
- Complexity and size of dataset
- Meaningful metadata not present

Challenges to Data Sharing...

On the librarian side ...

- Institutional support missing
- Skill-set missing
- Not knowledgeable about faculty research
- Not interested or knowledgeable about a subdiscipline
- Full work load already
- Reference/Bibliographer Model
- Cannot attend training off-campus
- New model of service helping in "publishing" research
- Lack of experience in data stewardship or data curation
- ► Economic constraints of institution who's going to pay for archiving & access?

Domain Repositories

Current mandates for researchers to manage data have one drawback – the false assumption that it will be cost-free, or that the state will automatically support it.

Hence, 25 discipline-based or domain repositories, meeting in Ann Arbor at ICPSR in 2013, drafted an open letter urging the United States Government to commit funds to support managing of research data and its preservation and accessibility.

As libraries move to implementing data management services it is important to acknowledge the pioneering efforts of such repositories as ICPSR, social science data archives (including the Australian Data Archive at ANU) & its data archivists, and professional organizations like IASSIST which have been playing a leading role archiving and disseminating datasets for up to half a century.

ICPSR: http://www.icpsr.umich.edu/

ADA: http://www.ada.edu.au

IASSIST: http://www.iassistdata.org

Challenge for Domain Repositories

- Despite the growing demand for data sharing and access, domain repositories face an uncertain financial future in the United States. The need for data archives is rising due to open access mandates, research innovations, and the growing volume of scientific data that needs to be curated, preserved, and disseminated. Yet funding for domain repositories remains unpredictable and inadequate for the task at hand. Of particular concern is the mismatch between the long-term commitments to preservation inherent in the work of archiving, and the short-term and episodic funding upon which this work is based. Many archives rely primarily on project-based grants, even though the expectation of stakeholders is that data will be available and usable indefinitely.
- Another concern is that the push towards open access, while creating more equity of access for the community of users, creates more of a burden for domain repositories because it narrows their funding possibilities. Without care, this may create a different kind of inequity-- less well-funded scholars or institutions will be less likely to have their products of research preserved for the future.

Domain Repositories' Call For Change

A Call for Change

- Domain repositories must be funded as the essential piece of the U.S. research infrastructure that they are. This means: Ensuring funding streams that are long-term, uninterrupted, and flexible
- Creating systems that promote good scientific practice
- Assuring equity in participation and access

There may not be one solution to the problem. Repositories may very well need different funding models across domain and repository type. But in every case, creating sustainable funding streams will require the coordinated response of multiple stakeholders in the scientific, archival, academic, funding, and policy communities.

Source: "Sustaining Domain Repositories for Digital Data: A Call for Change from an Interdisciplinary Working Group of Domain Repositories" - http://tinyurl.com/domainrepositories25

Benefits to Sharing Data

- Contributing to scientific knowledge
- Potential of new discoveries & understanding based on secondary analysis
- Researchers get cited more often if their data is "published"
- Publicly funded research made transparent
- The public benefits from open access to data
- Scientific fraud risk diminished
- Improves profile of institution
- ► Enhances scholarly collaboration & communication
- Repurposes the library mission for the foreseeable future

Online Courses

University of Edinburgh offers an excellent array on self-paced online instruction on a variety of topics:

- Research data explained
- Data management plans
- Organising data
- ► <u>File formats & transformation</u>
- Documentation & metadata
- Storage & security
- Data protection, rights & access
- Sharing, preservation & licensing
- Software practicals

Source: http://datalib.edina.ac.uk/mantra/

DIY Training Kit for Librarians (Edinburgh University)

- Promotional slides for the RDM Training Kit
- Training schedule
- Research Data MANTRA online course by EDINA and Data Library, University of Edinburgh
- Reflective writing questions
- Selected group exercises (with answers) from UK Data Archive, University of Essex
- Podcasts for short talks by the original Edinburgh speakers if running course without 'live' speakers
- Evaluation forms
- Independent study assignment: Interview with a researcher, based on Data Curation Profile

Source: http://datalib.edina.ac.uk/mantra/libtraining.html

Data Management for Clinical Research (Online course, Vanderbilt University)

"This course is designed to teach important concepts related to research data planning, collection, storage and dissemination. Instructors will offer information and best-practice guidelines for 1) investigator-initiated & sponsored research studies, 2) single- & multi-center studies, and 3) prospective data collection & secondary-reuse of clinical data for purposes of research. The curriculum will balance theoretical guidelines with the use of practical tools designed to assist in planning and conducting research. Real-world research examples, problem solving exercises and hands-on training will ensure students are comfortable with all concepts."

Source: https://www.coursera.org/course/datamanagement

Acknowledgements

- ► Thanks to all the online sources from which I have included material for this presentation.
- ► And my gratititude to SMU Library, especially Gulcin Cribb for inviting me to speak and Yeo Pin Pin for making the arrangements.

Thank you!

Source: 資深圖書館員曾振鍛獲獎, Chinese Daily News, 18 August 2013



資深圖書館員曾振鍛獲獎

【爾灣訊】

August 18, 2013 09:00 AM | 3844 次 | 0 🖳 | 13 🎍 | 🖂 | 📥



爾灣加大圖書館資深館員曾振鍛,榮獲美國政治與社會科學研究組織傑出服務獎。(爾灣加大提供)

美國政治與社會科學研究組織是美國規模最大的社會科學資料庫,該組織每兩年選出一位傑出服務獎得主,並在該會年會中頒獎表揚,過去16位傑出服務獎得主包括任職耶魯大學、麻省理工學院及柏克萊加大等名校的學者。曾振鍛自1986年起即擔任該組織在爾灣加大的代表,為南加州地區的社會科學研究學者及學生提供有關如何使用該資料庫的協助。

來自香港的曾振鍛獲有密西根大學圖書館碩士,他在爾灣加大負責總圖書館有關政治、經濟與亞美研究的資料庫,他曾於2004年獲選為傅爾布萊特學者(Fulbright Scholar),並前往越南實地了解當地的圖書及研究資料庫,他也發表過許多與越南難民及美國越南裔社區相關的文章。