

Patron Driven Data Acquisitions: Prizes, Perils, and Pitfalls



Jim Church & Josh Quan
Beyond the Numbers: The Economic Data Ecosystem
Federal Reserve Bank of St. Louis, November 2018

Background

- Faculty and students requesting data sets from librarians for years, particularly in economics.
- Requests for specialized data sets (robots, gasoline, cryptocurrencies, rental prices).
- Requests for restricted or “sensitive” data requiring secure storage and research review.
- Data is expensive and cost prohibitive for most graduates and undergraduates.
- No data, no research. You can’t interlibrary loan a data set.

The Assumptions

Faculty and graduate students typically know what data they need.

As professional researchers, often better than librarians do.

They generally do not approach librarians until other options are exhausted.

The Pre-DAAP Landscape

- Students and faculty sent librarians requests for data.
- Librarians scrambled and horse-traded to get the funds.
- Swaps and deals. Joint funding with departments.
- Special one-time funding debates at selector meetings.

Questions

- Why should libraries collect data sets?
- Don't we have it already? Did you look in...?
- Shouldn't all data be open access?
- Why should we spend money on just a few researchers?

Aspirations


“James Church (2008) describes the challenges of acquiring international survey microdata, much of which must be purchased and is in high demand by researchers, and recommends strategic data acquisition in consultation with researchers. He advocates for a coordinated, user-driven acquisitions model that includes data within its purview, stating *“we buy many books, documents, and microfiche that few people use...it seems misguided to spend thousands of dollars buying items students have not specifically requested while neglecting to purchase data that could lead to the publication of a doctoral thesis”* (p. 16). Church describes one particular type of data and user need, but these principles can be extended to many disciplines and data types.”

-- “Assessing a Patron-Driven, Library-Funded Data Purchase Program.” *The Journal of academic librarianship* Sheehan, Beth yr:2017 vol:43 iss:1 pg:49.

The Road to DAPP

- Librarians and campus D-Lab partners make the case for acquiring restricted use data and patron initiated data requests.
- Economics grad students distribute letter and a petition.
- Director of the D-Lab liaises with the Library.
- The library forms the first of many data working groups.
- A central data fund set up for one time purchases.

The Illinois Data Purchase Program



Scholarly Commons

306 Main Library | 1408 West Gregory Drive | Urbana IL 61801 | 217-244-1331 | sc@library.illinois.edu

[Home](#) [About Us](#) [Services](#) [Guides](#) [Workshops](#) [Events](#)

[Scholarly Commons](#) > [Data Purchase Program](#)

Data Purchase Program

Hours
CURRENTLY:
Today 8:30 AM - 6:00 PM

[more](#)

Program description

The University Library is soliciting applications from faculty, graduate students, and academic professionals who need to purchase numeric, spatial, or textual data for their research. The University Library will make the data purchased or licensed as part of this program available for general use by the campus community. Award amounts will vary depending upon need and the availability of funding from other sources such as departmental, college, or other grant funds. However, the availability of outside funding is not a condition of receipt of funding from the Library.

Awards will be targeted toward meeting smaller needs (ca. \$5,000 range). However, the amount available to be awarded for individual proposals will depend up on the total number and suitability of applications received.

Awards will be made on a competitive basis after the beginning of the fiscal year (July 1st), with orders being placed largely during the fall semester and with anticipated availability in the spring semester. Applications will be considered after the deadline as long as funds are available and the purchase can be completed by the end of the fiscal year (June 30th).

Eligible Researchers

Faculty, academic professionals, and graduate students at the Urbana-Champaign campus whose research requires numeric (i.e., survey, observational), spatial, or textual data.

Eligible Data Sets

Data requests can be for repackaged public data or for data collected by individual researchers that is subsequently made available for a fee. The library strongly prefers to acquire data that are accessible online rather than on DVD or other tangible storage media. For licensed data, this grant program will only support a license for one year. Other library funds may be available to continue the license after the first year but continued access cannot be guaranteed. Library funds will only be used to purchase data that can be made available to researchers across campus.

Library/D-Lab Working Groups



Data Sharing, Storage and the Social Sciences Working Group

Recommendations and Report

December 9th, 2016

Working Group Membership:

Jamie Wittenberg (Chair)

Jim Church

Harrison Dekker

Celia Emmelhainz

Jon Stiles

Data Purchase Working Group

PRELIMINARY REPORT

FEBRUARY 2015

HARRISON DEKKER, CHAIR

JIM CHURCH

LYNNE GRIGSBY

LYDIA PETERSEN

SUSAN POWELL

TRINA PUNDURS

HILARY SCHIRALDI

ELLIOTT SMITH

D-Lab/Library Restricted Use Data Partnership

The screenshot shows the D-Lab website interface. At the top is a dark blue header with the D-Lab logo and the tagline "Intelligent research design for data intensive social science". Navigation links include ABOUT, SERVICES, RESOURCES, BLOG, PROJECTS, CALENDAR, and SEARCH. Below the header is an orange bar with the word "Data". A link to "Sign up for our mailing list" is present. On the left, a "RESOURCES" sidebar lists "Data Resources", "Campus Resources", and "Course List". The main content area contains several paragraphs of text explaining the data acquisition initiative and access restrictions. A list of resources follows, including the "India - National Sample Survey - Consumer Expenditure Survey Series". A detailed view of this resource is shown at the bottom, including its title, a link to the data use agreement, and a description of the survey.

Intelligent research design for data intensive social science

ABOUT SERVICES RESOURCES BLOG PROJECTS CALENDAR SEARCH

Data

Sign up for our mailing list

RESOURCES

- ✓ Data Resources
- ✓ Campus Resources
- ✓ Course List

THESE DATA WERE ACQUIRED BY THE LIBRARY AS PART A NEW DATA ACQUISITION INITIATIVE, PILOTED IN SPRING 2016, AND ARE JOINTLY HOSTED AND MANAGED BY THE D-LAB AS A PARTNER IN THAT INITIATIVE.

THESE DATA ALL HAVE RESTRICTIONS ON ACCESS, USE AND DISTRIBUTION OF THE RESOURCES.

ACCESS CAN BE PROVIDED TO BERKELEY AFFILIATES WHO AGREE TO COMPLY WITH THE RELEVANT DATA USE AGREEMENT. TO APPLY FOR ACCESS, COMPLETE THE SURVEY LINKED TO UNDER THE DATA USE AGREEMENT PORTION FOR EACH RESOURCE.

India - National Sample Survey - Consumer Expenditure Survey Series
India Annual Survey of Industries, 2008-2013
San Francisco Chronicle Archive
Los Angeles Sentinel Archive
Airbnb data for Six Metro areas, 2014-2016
NYSE ReTrac ProTrac EOD, 2007-2016
Corpus of Contemporary American English (COCA)
Catalist

Title: **India - National Sample Survey - Consumer Expenditure Survey Series**

Data Use Agreement: < [Apply for Access Here](#) >

Description: Surveys of expenditures incurred by households on the consumption of goods and service during the reference period, as well as enterprises owned by households and used by their members during the reference period. Similar surveys are used in the United States and other countries and are used to track consumption and spending patterns of the population and growth patterns for consumer goods.

Planning Considerations

- Eligibility
- Licensing/Acquisition
- Systems Support
- Discovery & Cataloging
- Storage
- Publicity

Initial Eligibility Considerations

Numeric, geospatial, and textual data will all be eligible. Funding for data memberships and for software in support of data services will also be supported.

Restricted use data sets and data sets available to only select campus users are eligible for funding. Priority will be given to data sets available to all UCB students, faculty, and staff.

Cost sharing is encouraged, and may increase the likelihood of a proposal being funded.

Grants awarded quarterly and limited to \$5000. Emphasis on multi-disciplinary data sets.

Publications must contain the following acknowledgment: "The author wishes to acknowledge the Library at the University of California, Berkeley, which provided support for the completion of this research."

Eligibility in Practice

Due to demand, grants were eventually awarded on an ongoing basis.

Regular meetings to consider grants were abandoned.

The initial cost limit (\$5000) was an underestimate. The average cost was over \$8000 and several data sets were considerably more.

Funding was transferred from central collections to the data services librarian.

Acquisitions

- Typical Scenario - Grad Student or faculty email a librarian to inquire about a data source.
- Librarian meets with the researcher to ascertain eligibility, costs, and relevance.
- Potential cost sharing options discussed with colleagues and departments.
- Librarian and/or graduate student contacts the vendor to make inquiries.
- Librarian Refers researcher to the [Data Acquisition and Access Program web page](#).
- Researcher submits the proposal using the [Library Purchase Recommendation Form](#).

Data Acquisitions Program Web Site

BERKELEY

Library

UNIVERSITY OF CALIFORNIA

Library / Library Guides / Data Acquisition and Access Program / UC Berkeley Library

Data Acquisition and Access Program: UC Berkeley Library

The Data Purchase Program invites applications from UCB researchers in all disciplines.

UC Berkeley Library

FAQ

About

The Data Acquisition and Access Program provides a method for UC Berkeley faculty, students and staff to get access to data (numeric or textual) that requires purchase, licensing or is otherwise restricted. Made possible through a partnership between the [University Libraries](#) and the [D-Lab](#), this program is focused on datasets that require license or user agreements to access.

The program has an open call for faculty, students and staff to submit requests for data. More information on the criteria and eligibility for inclusion in this program are below.

Criteria and Eligibility

The University of California Berkeley Library is accepting applications for a program to acquire data sets (through purchase or licensing) to support campus research needs. The Library prefers to acquire data sets that are accessible online, rather than on tangible storage media, and prefers products that contain software-independent formats. In order to maximize impact, the data sets must be of interest to more than one research group or department. If the data set is licensed on a subscription basis, only one year will be paid for by this program.

Eligibility: All UC Berkeley faculty, students, and academic researchers (or UCB librarians on their behalf) are eligible to apply; undergraduate requests must be supported by a faculty member.

Terms of Use: The data set must be accessible to all Berkeley faculty, students, and staff. In addition, the Library and/or D-Lab must be able to comply with the requirements and/or restrictions associated with data and the terms must be in accordance with the Library's standards for licensing agreements.

Documentation: Documentation describing the data set and how to access and use it must accompany the data set.

Other Requirements: Researchers who receive funding may be asked to give a library-sponsored informal talk about how they are using the data, focusing on methodology and issues related to the data rather than on the results of their research. Publications based on the data sets purchased with these funds must contain the following acknowledgment: "The author wishes to acknowledge the Library at the University of California, Berkeley, which provided

5358973208462643829795208418714839937316582074094592307014040520630099802001
0451328208647093844609550582221725359081284411174507841027019385211055596444
09756659334461284756482337867831652712019091456485669234603486104543264821"
0066063158817488150920961629254091713463678925703600113303305488204665"
7026575959195209218617280193261178310511854007465227994274956025108595272"
62440656643086021394946395264271907021"360943202770539717162937625"
714526356082778577134275778960917363717"68440012249534301465493"
6112150216086403441015911362977407130"0721134999999832970"
5336400830261952200253344640332619"0133483837086580"
042875546873115956286388235378759375"3217122680661"
01065485863278665936153381827968230"89957736"
7454158695950829533118687272855894"3192"
5637076610471018194295559619894667"
28989152104752162056966740580381"
1636009341721641219924"
182202255402648044"
161452491921321721"
62518983102544"
0141


DATA


Submit Your Request!

To apply for funding, please select **Data Set** on the [Purchase Request Form](#) and fill out the required elements.

Purchase Recommendation Form

[ASK US](#) | [MY OSKICAT](#) | [HOURS AND MAPS](#) | [CONNECT FROM OFF CAMPUS](#) | [UC BERKELEY HOME](#)





[HOME](#) | [LIBRARIES](#) | [USING THE LIBRARIES](#) | [RESEARCH SUPPORT](#) | [ABOUT](#) | [HELP](#)

Using The Libraries

- List of libraries
- Borrow
- Renew
- Pay fines
- Connect from off campus
- Find a study space
- Reserve a study room
- Interlibrary Borrowing and Lending
- Print/scan
- Have a librarian meet with your class
- Computers in the libraries

Purchase Recommendation Form

[Questions?](#) [Ask Us!](#)
Chat • Phone • Email
Find a subject librarian

- The Library welcomes suggestions for titles to consider purchasing.
- Your suggestion will be evaluated based upon the Library's collection development policies and current budgetary constraints.
- Please provide as much information about the suggested title as you have available.
- In addition to submitting this purchase recommendation, you may be able to borrow a copy of this work through the [Interlibrary Borrowing Service](#).

Item Information

Type of material (required): *

Subject (required): *

Title (required): *

Author:

Materials Acquired Digitally (MAD)

Electronic Resource Order Request Form (purchased and remotely hosted)	
Please provide as much of the following information as possible. An asterisk (*) denotes a required field	
Name: James A. CHURCH Email: jchurch@library.berkeley.edu Campus Phone: +1 510 768-7611	
*1. Title of resource:	<input type="text"/>
*2. Publisher/Provider:	<input type="text"/>
*3. Publisher Contact Name:	<input type="text"/>
*4. Type of resource (Select one; enter details in the "Other" box below):	
<i>E-journal (single e-journal title)</i>	
<input type="radio"/> Online only	
<input type="radio"/> Online only, cancel existing print order	
<input type="radio"/> Add online to existing print order	
<input type="radio"/> New print + online	
<i>E-journal package (several e-journal titles)</i>	
<input type="radio"/> E-journal package	
<input type="radio"/> E-journal package, with cancellation of print titles. <i>If applicable, include details in the "Other" box below</i>	
<i>Journal backfile/archive</i>	
<input type="radio"/> Single title backfile	
<input type="radio"/> Backfile of multiple titles	
<i>Ebook(s)</i>	
<input type="radio"/> Single title	
<input type="radio"/> Collection or package (multiple titles)	
<i>Other</i>	
<input type="radio"/> Please describe below	
<i>Other resource type information</i>	
<input type="text"/>	

Vendor Negotiations

- This can be time consuming - weeks and even longer are not unusual.
- Some commercial data vendors have little to no experience with libraries.
- Licensing may be inappropriately restricted (or relaxed). Initial review of licenses by the librarian is important.
- Negotiations may involve differences in communication styles and legal environments.

Licensing

Data licensing can also be both complex and time consuming.

Many data providers are unaccustomed to dealing with libraries:

- Data restricted by department
- Data only for faculty and Ph.Ds
- Data requiring a signature and account for each user
- Data that needs to be returned/destroyed after license expires

Our licensing librarian worked for extended periods on some licenses.



Ticketing

“Every stage of the electronic resource life cycle, from trial, to acquisition, to activation, to maintenance and troubleshooting, to renewal or cancelation, requires a complex series of nonlinear tasks, typically involving hand-offs between multiple library staff members and follow-ups with external players such as students, faculty members, administrative offices on campus, and/or vendors.”

“Librarians and staff nevertheless must find ways to complete the work. Often, we struggle through our efforts using a combination of emails and shared spreadsheets. This can often lead to communication issues due to a lack of transparency, poor service outcomes and frustration as issues reported via emails get buried in staff inboxes, confusion as to whom has completed what tasks, and lack of documentation.”

-- Smith, Kelly. “Problem Solved! Managing Electronic Resource Workflows Using Ticketing System Software,” *Serials Review*, V42:1, (2016). <http://dx.doi.org/10.1080/00987913.2015.1137674>

Graduate Students

Careers and/or degrees can be determined by the success or failure of a data project.

Grad students & faculty have domain expertise to identify needs and issues.

Use caution in allowing grad students to interact with vendors.

Ensure students understand usage provisions, e.g. not taking data with them after they graduate or sharing restricted data.

Undergraduates

Some of the most successful research at UC Berkeley conducted by undergraduates.

Limiting proposals to faculty & grad students raises equity questions.

Undergraduates have shorter time constraints. Make sure they understand failure is a possibility and the process can take time.

Selector Due Diligence

Don't forget documentation

Don't forget about file sizes & formats

Don't forget to get data samples

Consider all aspects of confidentiality

Anticipate technical/computing constraints



Systems Support

- File sizes, data structures, and transfer speeds are considerations that should be addressed by selectors & liaisons.
- IT not equipped for making large (>50Gb) datasets available to researchers easily.
- Complicated further by large hierarchical datasets of varying quality (XML).
- The responsibility for understanding the data and what is expected upon delivery falls to the selector and requesting academic.

Ticketing and Communications

- Current workflow uses Footprints, a proprietary + legacy system designed for tracking events during the acquisitions process.
- CORAL, an open-source electronic resource management system, is being investigated as a supplemental tool to manage complex licenses, communication, and workflows.

Expectations and Time Constraints

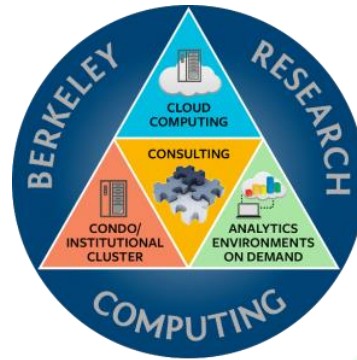
Some data can take weeks or months to process. The longest took several months.

Lost checks, administrative delays, challenging licenses, communications issues.

Multiple parties involved in a nonlinear process: licensing, acquisitions, IT and catalogers.

Students and faculty need to have realistic time expectations.

Computing and Storage Environments



Plans for a Berkeley Dataverse



Cataloging and Metadata

- Data cataloging is challenging.
- Complex workflow issues.
- Lack of catalog records for unique data sets.
- Neither the cataloger nor selector permitted to view restricted or sensitive data.

Sample Metadata

The screenshot shows the OskiCat UCB Library Catalog interface. At the top, there's a navigation bar with links like 'Start Over', 'Help', 'What's Included', and 'Library Home'. Below this is a search bar with various filters and buttons like 'START OVER', 'ADD TO MY LISTS', 'SAVE RECORDS', etc. The search results show 36 items found, sorted by date. The first result is 'NYSE ReTrac ProTrac EOD, 2007-2015' by NYSE. The description mentions it's a 1 online resource containing trading data from the New York Stock Exchange. A callout box points to the 'MAD Dataset Record CU' link in the 'Added Title' field, with the text 'Materials Acquired Digitally'.

Uniform Title [NYSE ReTrac ProTrac EOD, 2007-2015.](#)

Title **NYSE.**

Published [New York, New York] : [New York Stock Exchange, LLC], [2007-2015]
[Berkeley, California] : [University of California, Berkeley, Library], 2016.

Link to online version(s):

[Request access to download dataset.](#)

Description 1 online resource

Summary Trading data from the New York Stock Exchange for every business day 2007 through 2015, as follows: NYSE ProTrac EOD Replay (in folder "TRADE_PROGRAM") contains a summary of all program trading activity for each stock identifying the total amount of shares that was executed for each stock through index arbitrage program trading vs. all other program trading strategies. NYSE ProTrac EOD Summary (in folder "SUM_PROGRAM") contains all program trading activity for each stock identifying the total amount of shares that was executed for each stock through index arbitrage program trading vs. all other program trading strategies. NYSE ReTrac EOD Replay (in folder "TRADE_RETAIL") is a record-by-record replay of the ReTrac datafeed, containing the Symbol, Volume, and Time of each retail execution. NYSE ReTrac EOD Summary (in folder "SUM_RETAIL") provides an end-of-day summary of all retail trading activity for each stock identifying the total amount of shares that were bought vs. sold for each stock that was traded by retail customers; shares are totaled for each symbol in separate buy and sell quantities.

Subject [New York Stock Exchange -- Statistics -- Databases.](#)
[Stocks -- United States -- Statistics -- Databases.](#)

Added Author [New York Stock Exchange. Issuing body.](#)

Added Title **MAD Dataset** [Record CU](#)

Direct Link <http://oskiat.berkeley.edu/record=b23650737~S1>

ALSO LOOK IN

- UC LIBRARIES AND BEYOND
- STANFORD
- Google books

Materials Acquired Digitally

Restrictions on Use

- As required by U.S or National Law (US Census Bureau, EU, Statistics Canada, UK Data Center).
- As stipulated by the vendor.
- As required by the University (Institutional Research Board, Human Subjects).

When in doubt err on the side of caution and consult a licensing or legal expert

Qualtrics & Terms of Use



NYSE EOD summary data, 2007-2016

Need more information about these data?

See: http://dlab.berkeley.edu/data-resources/data#NYSE_EOD

TERMS OF USE

You are applying for access as an authorized user to extract and/or compile data from our locally loaded copies of the archived licensed content, provided that all such activities are within the bounds for fair use and are for the teaching, learning and research purposes. All supplemental terms, including any applicable display and use restrictions continue to apply.

Please read the terms of use below. If you agree to them, click on the "I Agree" button to proceed. If you do not agree, you can click on the "I Do Not Agree" button to return to the home page.

I agree that I will use these datasets solely for non-commercial research, statistical, or educational purposes and NOT for investigation of specific research subjects or commercial gain (i.e. by reselling, redistributing or republishing licensed content).

I agree that I will NOT use the materials in any way prohibited by applicable laws

I agree

I do NOT Agree

I agree that my access to these data is contingent on my active affiliation as faculty, staff or student with UC Berkeley, and I will cease to use these data when my affiliation ends.

I agree

I do NOT agree

I understand that some uses of the data may require approval of the [Berkeley Office of Human Subjects](#). I have looked at the [guidelines](#) and (select one option below)

My use of the data does not require IRB approval

I have obtained IRB approval for my use of the data

I need to obtain IRB approval, but have not done so yet

>>

Publicity

- ★ Reach out to faculty and students
- ★ Strike the right balance
- ★ Make time and cost constraints clear
- ★ Do not over-promise

Lessons Learned

- Some attempts to acquire data ended in failure, at times after long time periods (negotiations)
- Some data did not fully meet the needs of the researchers after the data arrived (due diligence)
- Some data so large (12 Tb) it took extended periods to download. Use was problematic due to storage and computing constraints (storage/computing)
- Some licenses took exceptionally long times to finalize (licensing)
- Some data had confidentiality issues for UC Berkeley even though these were not specified in the license (institutional restrictions)
- Some data was “lost” at various points in the process (workflows/ticketing)

Success Stories

Significant Research Data Sets were acquired from the program which offered new research opportunities to grad student and faculty, including:

- **India National Sample Survey Data, Consumer Expenditure Series (Seven Rounds) and Annual Survey of Industries (Four Rounds)**
- **Linguistic Data Consortium (292 data sets)**
- **Nielsen Consumer Panel and Retail Scanner Data**
- **AirBnb Data for Six Metropolitan Areas (2014-2016)**
- **Imports and Exports of Related Parties, US Census Bureau**
- **NYSE Retracs/Protracs data (Firm-level trading of shares by retail investors on the NYSE)**
- **Catalyst (voter registration data)**

AirBnb (AirDNA)

(Search History)

Result page: PREVIOUS NEXT

Title


Airbnb data.

Published

[Denver, Colorado] : [Airdna], [2016]
[Berkeley, California] : [University of California, Berkeley, Library], 2016.

Link to online version(s):

[Request access to download dataset.](#)

 **Description**

1 online resource (21 CSV text files)

Summary

Market summary reports of **Airbnb** listings and rental performance for six metropolitan areas in the United States, including revenues and property features.
Chicago-Naperville-Elgin, IL-IN-WI MSA: daily, August 2014-September 2016; monthly, August 2014-September 2016; annual, year ending October 2016 -- Houston-The Woodlands-Sugar Land, TX MSA: daily, August 2014-September 2016; monthly, August 2014-September 2016; annual, year ending October 2016 -- Los Angeles-Long Beach-Anaheim, CA MSA: daily, August 2014-September 2016; monthly, August 2014-September 2016; annual, year ending October 2016 -- New York-Newark-Jersey City, NY-NJ-PA MSA: annual, year ending October 2016 -- Portland-Vancouver-Hillsboro, OR-WA MSA: daily, August 2014-September 2016; monthly, August 2014-September 2016; annual, year ending October 2016 -- San Francisco-Oakland-Hayward, CA MSA: daily, August 2014-September 2016; monthly, August 2014-September 2016; annual, year ending October 2016.

Subject

[Hospitality industry -- United States -- Statistics -- Databases.](#)
[Rental housing -- United States -- Statistics -- Databases.](#)
[Airbnb \(Firm\) -- Statistics -- Databases.](#)

Added Author

[Airdna \(Firm\), issuing body.](#)

Added Title

Title on resource download page: Airbnb
[MAD Dataset Record CU](#)

Direct Link

<http://oskicat.berkeley.edu/record=b23599036~S1>

India National Sample Surveys

Title: India - National Sample Survey - Consumer Expenditure Survey Series

Data Use Agreement: < [Apply for Access Here](#) >

Description: Surveys of expenditures incurred by households on the consumption of goods and service during the reference period, as well as enterprises owned by households and used by their members during the reference period. Similar surveys are used in the United States and other countries and are used to track consumption and spending patterns of the population and growth patterns for consumer goods.

Detailed metadata can be found at: The International Household Survey Network for:

(Schedule 1.0, Type 1) **1983, 1987-1988, 1993-1994, 1999-2000, 2004-2005, 2005-2006, 2006-2007, 2007-2008, 2009-2010, 2011-2012**

(Schedule 1.0, Type 2) **2009-2010, 2011-2012**

(Schedule 10.0) **1983, 1987-1988, 1993-1994, 1999-2000, 2004-2005, 2009-2010, 2011-2012**

Producers: Central Statistics Office (Industrial Statistics Wing) - Ministry of Statistics and PI, Government of India

Universe: Households and personal enterprises in India, with exceptions for inaccessible areas.

Geographic coverage: India. The survey covers the whole of the Indian Union except (i) interior villages of Nagaland situated beyond five kilometres of the bus route and (ii) villages in Andaman and Nicobar Islands which remain inaccessible throughout the year.

Time Period: 1983 - 2012, Assorted Years, Annual Surveys

Unit of Observation: Households

Smallest Geographic Unit:

Title: India Annual Survey of Industries, 2008-2013

Data Use Agreement: < [Apply for Access Here](#) >

Description: The Annual Survey of Industries (ASI) is the principal source of industrial statistics in India. It provides statistical information to assess changes in the growth, composition and structure of organised manufacturing sector comprising activities related to manufacturing processes, repair services, gas and water supply and cold storage. The survey is conducted annually.

Detailed metadata and documentation can be found at the International Household Survey Network for:

2009-2010, 2010-2011, 2011-2012, 2012-2013

Producers: Central Statistics Office (Industrial Statistics Wing) - Ministry of Statistics and PI, Government of India
Citation: Smallest Geographic Unit:

Geographic Coverage: India. The geographical coverage of the Annual Survey of Industries, 2008-2009 has been extended to the entire country except the states of Arunachal Pradesh, Mizoram and Sikkim and Union Territory of Lakshadweep.

Time Period: 2009 - 2013, Annual

Unit of Observation: The unit of enumeration is the establishment - a factory, workshop, or undertaking - although a consolidated return is permitted for establishments in the same state and industry that share common ownership.

Universe: The survey cover factories registered under Sections 2m(i) and 2m(ii) of the Factories Act, 1948 i.e. those factories employing 10 or more workers using power; and those employing 20 or more workers without using power. The survey also covers bidi and cigar manufacturing establishments registered under the Bidi & Cigar Workers (Conditions of Employment) Act, 1966 with coverage as above. All electricity undertakings engaged in generation, transmission and distribution of electricity registered with the Central Electricity Authority (CEA) were covered under ASI irrespective of their employment size. Certain servicing units and activities like water supply, cold storage, repairing of motor vehicles and other consumer durables like watches etc. are covered under the Survey. Though servicing industries like motion picture production, personal services like laundry services, job dyeing, etc. are covered under the Survey but data are not tabulated, as these industries do not fall under the scope of industrial sector defined by the United Nations. Data Types:

Catalist

Title: Catalist


Catalist contains data from more than 240 million unique voting-age individuals from all 50 states and the District of Columbia. Data is compiled from a number of sources, including voter registration and Census data. Catalist provides two access points: the Q Tool and the M Tool. The Q Tool is a query to conduct extracts of the database, and the M Tool is used for uploading the researcher's data to the database and matching voter records from the researcher's dataset to the database information.

IMPORTANT: Use is limited to the UCB community only. All UCB users must apply for access via the D-Lab through the linked form. Staff at the D-Lab will contact the user when access is approved.

Data Use Agreement: < [Apply for Access Here](#) >

Nielsen

[ASK US](#) | [MY OSKICAT](#) | [HOURS AND MAPS](#) | [CONNECT FROM OFF CAMPUS](#) | [UC BERKELEY HOME](#)



Library
UNIVERSITY OF CALIFORNIA

[Library](#) / [Library Guides](#) / [Doe Library](#) / [Economics](#) / Nielsen Data

Economics: Nielsen Data

[Intro](#) [Articles & Working Papers](#) [E-books](#) [Data](#) [News](#) [Agricultural](#) [Behavioral](#) [Development](#) [Economic Demography](#) [Economic History](#)

[Environmental](#) [Finance](#) [Health](#) [Housing](#) [Labor](#) [International](#) [Law](#) [Political Economy](#) [Public Economics](#) [Nielsen Data](#)

Nielsen Registration Information

UC Berkeley subscribes to the Nielsen marketing data from Marketing Data Center at Chicago Booth. The James M. Kilts Center for Marketing at Chicago Booth and the Nielsen Company have partnered to make two consumer marketing datasets available to US-based academic researchers.

The Consumer Panel Data include longitudinal data beginning in 2004 with annual updates. These data track a panel of 40,000 - 60,000 US households and their purchases of fast-moving consumer goods from a wide range of retail outlets across all US markets. Retail Scanner Data consist of weekly pricing, volume, and store environment information generated by point-of-sale systems from more than 90 participating retail chains across all US markets. Data begin in 2006 and include annual updates.

UCB's campus agreement has the following restrictions:

- For PhD students who plan to use the data, *the student and the advisor are each required to register.*
- *A student in an undergraduate or masters program cannot be a registered user under this Agreement with direct access to the Data; access for such students must be made through a faculty member. Researcher who is registered with the Marketing Data Center and who indicates plans for directing and accepting direct responsibility for the student's access in a classroom or independent study setting.*
- Researchers will not transfer, sell, publish, redistribute or release or otherwise make available the Data, or any data contained therein, to any individual or third party who is not an employee, a faculty or staff member, or a student of UCB, and that it will not re-transmit the same from its institution for any purpose, including re-transmission to another premises operated by UCB without written permission from Chicago Booth.
- The Data may not be installed on personal systems (not institutionally owned) of faculty, staff, or students of UCB, although insubstantial extracts of data required for particular analysis processes may be temporarily downloaded to such systems as required for analysis and coursework, provided that steps are taken to assure that this data are purged once it is no longer required and/or the individual leaves the UCB.
- Data are provided on a "license to use" basis for a specific term, and that this license does not allow their use for private consulting or carry over to an individual's engagement in activities for another entity or institution, or if the individual should leave UCB, or if the license provided under the Agreement expires and is not renewed.
- The license to Data also does not carry over to or allow access and/or use by "co-authors" or "co-researchers" at other institutions, if the individual's institution is not also a Licensee to the necessary Data.
- This license does not permit any users, with particular attention to users such as visiting and adjunct professors, or students engaged in part-time employment or alternate research at another firm, consortium, or institution, to utilize the Data in these alternate endeavors, nor does this license carry over to any form of joint venture or consortium to which the subscribing institution belongs.

Conclusions

Acquiring and servicing data is like working with special collections. The material is expensive and requires additional effort to acquire, host, describe and service.

Data selectors need a knowledge of licensing, negotiations, file structures, and RDM.

Support from licensing, metadata specialists and systems administrators is essential.

Investing in a robust ticketing system potentially solves workflow issues.

The Wherefores

So undergraduates can write theses.

So grad students can write dissertations.

So faculty can write books and articles.

So librarians can support research teaching and learning

Making the World a Better Place



References

Church, Jim. "International survey data: Challenges and strategies for collection development." *DttP: A Quarterly Journal of Government Information Practice & Perspective*, 36 (1) (2008), p. 12-16.

Sheehan, Beth. "Assessing a Patron-Driven, Library-Funded Data Purchase Program." *The Journal of academic librarianship*. 43 (1), (2017), p 49-54.

Smith, Kelly. "Problem Solved! Managing Electronic Resource Workflows Using Ticketing System Software," *Serials Review*, V42:1, (2016), p. 59-64.

Image Credits

1. Slide 1 - "Pit." <https://en.wiktionary.org/wiki/pit>
2. Slide 8 - "Data Purchase Program." <https://www.library.illinois.edu/sc/purchase/>
3. Slide 19 - "Breach of Agreement" <http://picpedia.org/highway-signs/b/breach-of-agreement.html>
4. Slide 23 - "Baby Sledding" <http://www.afv.com/>
5. Slide 42 - Ashwini shankar. "Make the world smile by being conscious of your waste"
<https://i.pinimg.com/originals/65/81/11/6581111d1ad676dd8aa128c26a2d1578.jpg>