Data Acquisition and Access Program

“Joining the Library’s collection development strength with the D-Lab’s storage and auditing expertise”
daap pilot project team members

Jim Church
Harrison Dekker
Jennifer Dorner
Gail Ford
Lynne Grigsby
Mark Hemhauser
Erik Mitchell
Jean Mckenzie
Nga Ong
Trina Pundurs
Jon Stiles
What is *daap*?

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.
daap workflow

http://guides.lib.berkeley.edu/data
How does someone get access to a **daap** dataset?

**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 2) **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:** 2005 - 2012, Annual Surveys

**Unit of Observation:** Households

http://dlab.berkeley.edu/data-resources/data
<table>
<thead>
<tr>
<th>TITLE</th>
<th>IDENTITY</th>
<th>AmountPaid</th>
<th>Funding source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consumer Expenditure Survey.</td>
<td>India; Numeric data files</td>
<td>~$4000</td>
<td>DAAP</td>
</tr>
<tr>
<td>ProQuest Historical Newspapers the San Francisco chronicle.</td>
<td>DataSet - to be housed in the D-Lab</td>
<td>~$1000</td>
<td>DAAP</td>
</tr>
<tr>
<td>Annual Survey of Industries.</td>
<td>India; Rounds 19-22, on CD-ROM</td>
<td>~$8000.5</td>
<td>DAAP</td>
</tr>
<tr>
<td>AMADEUS [electronic resource].</td>
<td>2001-2006 on CD - One-time payment historical company data</td>
<td>~$8000</td>
<td>DAAP</td>
</tr>
<tr>
<td>Linguistic Data Consortium</td>
<td>To Begin: 2015-</td>
<td>~$7000</td>
<td>DAAP, linguistics dept</td>
</tr>
</tbody>
</table>
MOBILITY IN CITIES DATABASE 2015

“This database will be used to enhance the teaching of the class: Energy and Society (Energy and Resources 100 & 200, Public Policy C184 & C284). This is an interdisciplinary course that attracts hundreds students throughout the university.”

http://www.uitp.org/MCD2015
What’s next for daap?

1. Call for proposals open - funding dedicated for FY 2017
   i. NYSE ReTrac ProTrac EOD
   ii. Velocity Suite ("EV Energy" module)
   iii. MOBILITY IN CITIES DATABASE 2015

2. Exploration of appropriate discovery platforms

3. Marketing & outreach
Enabling better discovery
Unanswered questions

1. Is there a need across campus to put data that we have already acquired into this program?
2. Who should we reach out to in order to market the program?
3. How can/should we measure use of the data?
4. What kinds of data make sense to be included? Is there a use case for researcher generated data?