

# StatCat

## Building a Statistical Data Finder

[ssrs.yale.edu/statcat](http://ssrs.yale.edu/statcat)

*Steven Citron-Pousty*

*Ann Green*

*Julie Linden*

*Yale University*

# Themes

- Collaboration
- Domain-specific, not media or location specific
- Cross-media data finder
- Portal to Internet resources
- Numeric and spatial social science data

# Social Science Data Archive at Yale

- Digital collection since 1972
- Partnership between Social Science Library and Social Science Research Services
- Shared responsibility for the SSSA catalog

# History of the SSSDA Catalog

- Contained: Records for SSSDA holdings – data from ICPSR, Roper Center, federal agencies, IGOs/NGOs, commercial vendors.
- Designed as: SPIRES database on the mainframe, migrated to the Web.
- Maintained by: data librarian and Statlab

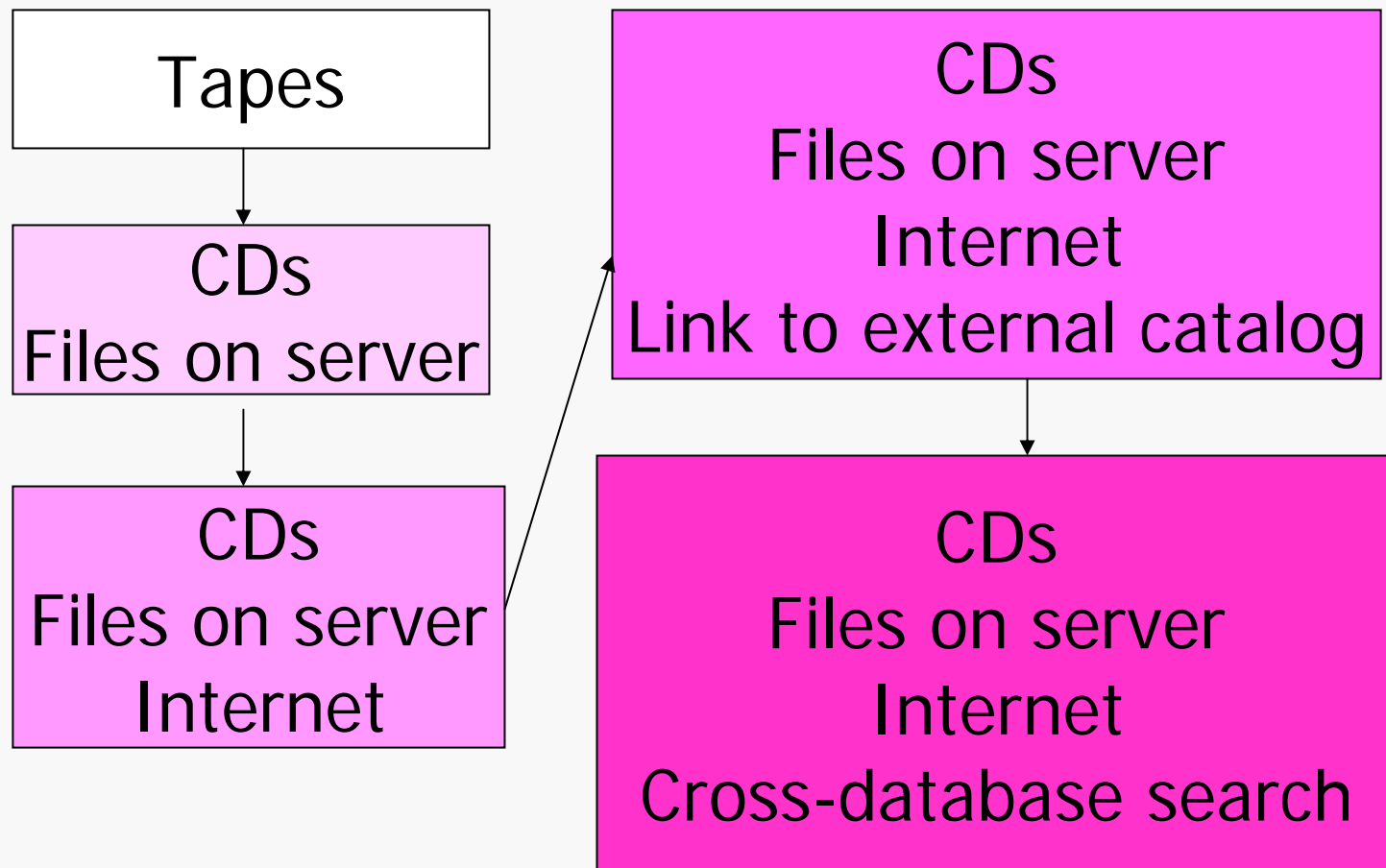
# The new catalog: StatCat

- Created a new structure to improve both front-end interface and back-end production and maintenance.
  - WAIS searching inadequate
  - Maintenance too difficult

# Goals for StatCat: domain

- Not a media-specific catalog, rather a domain-specific (social sciences) catalog.
- Includes datasets on Yale's Statlab server, CDs in the Library collections, and data available at other web sites.

# Evolution of StatCat



# Goals for StatCat: functionality

- Search fielded full text of records.
- Full location information to retrieve actual data.

# Goals for StatCat: Adhere to standards

- Base records upon a DDI subset (so that every field in StatCat maps to a DDI field).
- Potential output to multiple systems or metadata formats: MARC, DC, OAI, DDI, FGDC.

# Related Standards

<b>Bibliographic</b>	<b>Statistical Domain</b>	<b>Related domains</b>
MARC Dublin Core OAI GILS	DDI ISO 11179	FGDC EAD TEI

# Data Documentation Initiative

Consists of these parts:

- Document description
- **Study description**
- **File description**
- Data description
- Related material

# DDI Study Description section

- *Citation* – bibliographic information for the data collection
- *Scope* – information about the study's subject, geographic & temporal coverage (including abstracts and keywords)
- *Methodology & process* – information about how the data were collected (e.g. sample design)
- *Data access* – access conditions & terms of use for the data collection
- *Other study description materials*

# XML vs. Database

- XML is good at describing
  - 1) Hierarchical data
  - 2) Great for presenting multiple views into the same data source
  - 3) Exchanging data between independent sites in a highly structured manner
  - 4) Transport format: ASCII, fully tagged
  - 5) DDI and ICPSR are using it: will receive records in some version of DDI XML

# XML vs. database

- Decided to go with database and not XML at this time
  - Database met immediate requirements: improved searching and ease of maintenance. Well known technology.
  - XML tools still under development.
  - Drawback: records are no longer in “webspaces”
  - Eventually database will generate XML records.

# Designing the database

1. Determined what fields we needed:
  - Examined ICPSR's "slightly modified" version of the DDI codebook DTD and compared it to the current version of DDI.
  - Mapped our catalog fields to DDI.
  - Mapped out catalog fields to Dublin Core, looked at OAI.

```

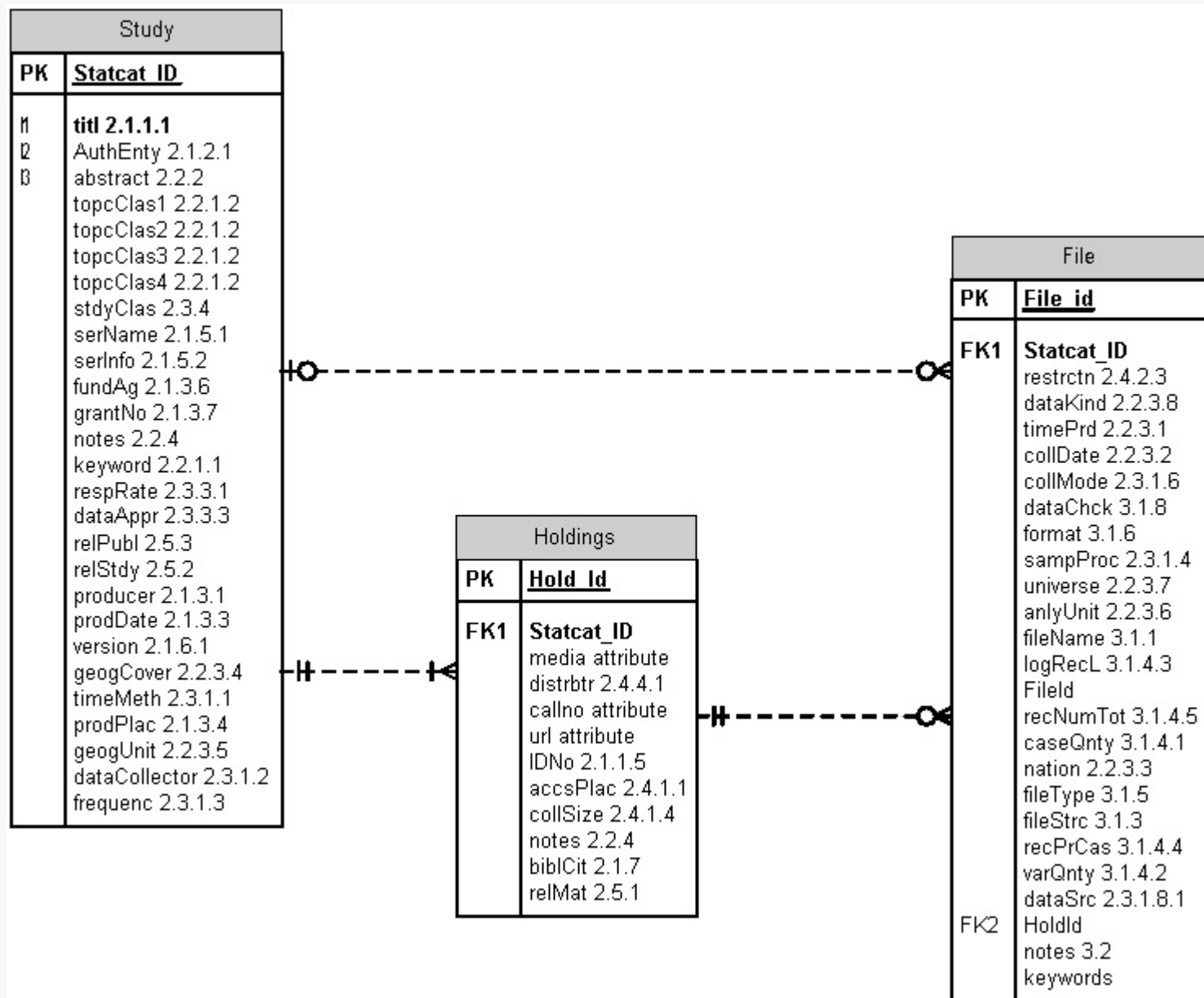
<?xml version="1.0" ?>
<?xml-stylesheet href="XSL/icpsr-abstract.xsl" type="text/xsl" ?>
<?xsl:output method="xslt" ?>
<?DOCTYPE codeBook
[
<codeBook>
<studyDscr>
  <citation>
    <titlStmt>
      <titl>[TITLE]</titl>
      <IDNo>[STUDYNO]</IDNo>
    </titlStmt>
    <rspStmt>
      <AuthEnty affiliation="[NIJ.AFFILIATION]">[INVESTIGATOR]</AuthEnty>
    </rspStmt>
    <prodStmt>
      <fundAg>[FUNDING.AGENCY]</fundAg>
      <grantNo>[GRANT.NUMBER]</grantNo>
    </prodStmt>
    <distStmt>
      <distrbtr abbr="ICPSR" affiliation="Institute for Social Research, University of Michigan"
=<http://www.icpsr.umich.edu">Inter-university Consortium for Political and Social Research</distrbtr>
      <distDate date="[DATE-ADDED]">[DATE-ADDED]</distDate>
    </distStmt>
    <serStmt>
      <serName>[SERIES.NAME]</serName>
      <serInfo>[SERIES.INFO]</serInfo>
    </serStmt>
    <verStmt>
      <notes>[COLLECT.CHANGES]</notes>
      <notes>[CHANGES.HISTORY]</notes>
      <dateAdded>[DATE-ADDED]</dateAdded>
      <dateUpdated>[DATE-UPDATED]</dateUpdated>
    </verStmt>
    <biblCit>[CITATION]</biblCit>
    <notes>[DATA.FORMAT]</notes>
  </citation>
  <studyInfo>
    <subject>
      <keyword>[SUBJECT.TERM]</keyword>
      <keyword>[NIJ.KEYWORDS]</keyword>
      <keyword>[KEYWORDS]</keyword>
      <topcClas source="archive" vocab="ICPSR subject classifications">[ICPSR.CLASSIF1]</topcClas>
      <topcClas source="archive" vocab="ICPSR subject classifications">[ICPSR.CLASSIF2]</topcClas>
      <topcClas source="archive" vocab="ICPSR Subject classifications">[ICPSR.CLASSIF3]</topcClas>
      <topcClas source="archive" vocab="ICPSR subject classifications">[ICPSR.CLASSIF4]</topcClas>
      <topcClas source="archive" vocab="HMCA subject classifications">[HMCA.CLASS1]</topcClas>
    </subject>
  </studyInfo>
</codeBook>
]

```

# Designing the database

1. Determined the type of queries we were going to ask of the data.
2. Determined relations between tables.
3. Determined which fields in which tables.

## StatCat database design (with DDI element numbers)



mission.

# Designing the database

4. Decided how to parse our records into the database fields.

# Side effects of the conversion process

- Scrutinize and clean up existing records
- Leads to questions: what are we cataloging, and why? What are we collecting, and why? Implications for archiving policies.

# StatCat v.2

- PHP migrated to a Java server-side application.
- More modular and extensible
- MySQL dbms migrated to PostgreSQL
- New avenues this opens
  - Spatial searches
  - Pre-analysis of data before downloading from our archive
  - Give the client metadata and data in the same download

# StatCat: Building a Statistical Data Finder

Search StatCat at Yale University - Microsoft Internet Explorer

File Edit View Favorites Tools Help

**StatCat** Yale University  
Social Science Statistical Laboratory  
Social Science Libraries and Information Services

*Search StatCat / Browse Subjects / Search Help / Help Using Data / About StatCat*

StatCat (formerly the Social Science Data Archive Catalog) includes information about numeric datasets in the Yale **Social Science Data Archive**, as well as datasets available in the Yale University Library and on the Internet.

**Search StatCat**

Enter text in one search field only

Keywords anywhere:   all words  phrase

ICPSR study number:

Field:    all words  phrase

**See also:**

- [ICPSR Direct](#): Faculty, students and staff on the Yale network can search the ICPSR (Inter-university Consortium for Political and Social Research) catalog and download data directly.
- [Yale Roper Collection](#): many of the historical Roper Collection datasets **are not** in StatCat.
- [Orbis](#), the Yale University Library online catalog, for datasets on CD-ROM and for codebooks that may not be included in StatCat.

**Browse Subjects**

<a href="#">Census Enumerations</a>	<a href="#">Health Care and Health Facilities</a>
<a href="#">Community and Urban Studies</a>	<a href="#">International Systems: Linkages, Relationships and Events Data</a>
<a href="#">Computer Programs and Instructional Packages</a>	<a href="#">Legal Systems</a>
<a href="#">Conflict, Aggression, Violence, Wars</a>	<a href="#">Legislative and Deliberative Bodies</a>
<a href="#">Economic Behavior and Attitudes</a>	<a href="#">Organizational Behavior</a>
<a href="#">Education</a>	<a href="#">Political Behavior and Attitudes</a>
<a href="#">Elites and Leadership</a>	<a href="#">Social Indicators</a>
<a href="#">Geography and Environment</a>	<a href="#">Social Institutions and Behavior</a>
<a href="#">Governmental Structures, Policies and Functions</a>	


The screenshot shows a Microsoft Internet Explorer browser window titled "Query Results in StatCat - Microsoft Internet Explorer". The address bar is empty. The menu bar includes "File", "Edit", "View", "Favorites", "Tools", and "Help". The main content area displays the StatCat website header with the Yale University logo and navigation links: "Search StatCat / Browse Subjects / Search Help / Help Using Data / About StatCat". Below the header, the search results are listed:

You did a title search on the words: county city data

- [American National Election Study, 1978, with Supplementary Data from the Congressional District Data Book and the County and City Databook](#)
- [County and City Data Book, 1972](#)
- [County and City Data Book, 1988](#)
- [County and City Data Book, 1994](#)
- [County and City Data Book \[United States\], 1983](#)

At the bottom of the page, there is a footer with contact information and a logo for the Data Documentation Initiative (DDI):

Yale University [Social Science Statistical Laboratory](#) and [Social Science Libraries and Information Services](#)  
Reference services: [Data & Electronic Services Librarian](#)  
Statistical consulting: [Statlab](#)  
This page last modified: Fri May 31 15:29:38 EDT 2002  
© 2001 Yale University

  
Metadata powered by the  
Data Documentation Initiative

StatCat Study Summary for Study #3018 - Microsoft Internet Explorer

File Edit View Favorites Tools Help

**StatCat** Yale University  
Social Science Statistical Laboratory  
Social Science Libraries and Information Services

[Search StatCat](#) / [Browse Subjects](#) / [Search Help](#) / [Help Using Data](#) / [About StatCat](#)

**Title:** County and City Data Book, 1994

**Author:** United States Department of Commerce. Bureau of the Census.

**Holdings available on:** [\[CD-ROM\]](#) [\[Internet\]](#) [\[All Holdings\]](#)

**Abstract:** Brings together a variety of social and economic data from the Census Bureau and other sources. Tables cover population, age, money and personal income, and education, as well as housing ownership and value. The report also includes births, deaths, poverty, local government finance and employment, health care, and human services. In addition, tables show data on business, banking, climate, elections, and social programs. Contains statistical tables for 194 data items for cities, and population and income data for smaller places. Statistics for all 3,141 U.S. counties, 1,078 cities with 25,000 or more inhabitants, and 11,097 places of 2,500 or more inhabitants. Includes socioeconomic and housing data from the 1990 census and the surveys that update them; city and county business statistics; median income, tax base, and more than 100 other variables for counties and cities nationwide.

**Subject hierarchy:** [All Subjects](#): [Census Enumerations](#): [Contemporary](#): [United States](#)

Holdings information - Microsoft Internet Explorer

File Edit View Favorites Tools Help

**StatCat** Yale University  
Social Science Statistical Laboratory  
Social Science Libraries and Information Services

*Search StatCat / Browse Subjects / Search Help / Help Using Data / About StatCat*

All Holdings for County and City Data Book, 1994

CD-ROM

[File details for this holding](#)

**Access location:** Mudd Library, Government Documents Center

**Call number:** C 3.134/2-1:


Internet

[File details for this holding](#)

**URL:** <http://fisher.lib.virginia.edu/ccdb/>

---

Yale University [Social Science Statistical Laboratory](#) and [Social Science Libraries and Information Services](#)  
Reference services: [Data & Electronic Services Librarian](#)  
Statistical consulting: [Statlab](#)  
This page last modified: Fri May 31 15:31:19 EDT 2002  
© 2001 Yale University

  
Metadata powered by the  
Data Documentation Initiative

## Holding Summary for Eurobarometer 52.0: European Parliament Elections, The Single European Currency, and Financial Services, October-November 1999 on CD-ROM

More current versions of this holding may be available on the ICPSR website.

Get this study [directly from ICPSR](#)

### [File details for this holding](#)

<b>Access location:</b>	Social Science Library Circulation Desk
<b>Call number:</b>	H 61.3 P47 PCD2000_004
<b>Distributor:</b>	ICPSR
<b>Distributor ID number:</b>	2892
<b>Files in this holding:</b>	1 data file + machine-readable documentation (PDF) + SPSS data definition statements
<b>Bibliographic citation:</b>	Melich, Anna. EUROBAROMETER 52.0: EUROPEAN PARLIAMENT ELECTIONS, THE SINGLE EUROPEAN CURRENCY, AND FINANCIAL SERVICES, OCTOBER-NOVEMBER 1999 [Computer file]. ICPSR version. Brussels, Belgium: INRA (Europe) [producer], 1999. Koeln, Germany: Zentralarchiv fuer Empirische Sozialforschung/Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributors], 2000.

# INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH

Social Science Data and Resources for Researchers



## Abstract--Study No. 2892

- [Download data set\(s\)](#)
- [View file information](#)
- [Printer-friendly page](#)

TITLE: Eurobarometer 52.0: European Parliament Elections, the Single European Currency, and Financial Services, October-November 1999

PRINCIPAL INVESTIGATOR(S): Melich, Anna.

### SUMMARY

This round of Eurobarometer surveys queried respondents on standard Eurobarometer measures, such as how satisfied they were with their present life, whether they attempted to persuade others close to them to share their views on subjects they held strong opinions about, whether they discussed political matters, how important the role of the European Union (EU) will be in their daily life over the next five years, and how they viewed the need for societal change. Additional questions focused on the respondents' knowledge of and opinions on the EU, including how well-informed they felt about the EU, what sources of information about the EU they used, whether their country had benefited from being an EU member, and the extent of their personal interest in EU matters. Other major areas of focus in the surveys included European Parliament elections, the single European currency (the euro), and respondents' experiences with financial services. Respondents were asked which television news programs they watched, which newspaper(s) they read, and how much attention they gave to politics, social issues, the European Union, the economy, sports, the environment, foreign policy, and culture. They were also asked whether they voted in the most recent European Parliament elections, what their reasons were for voting or not voting, and how they viewed the importance and power of the European Parliament in the European Union. A number of questions probed for respondents' knowledge and use of the euro, whether they worried about changing to the euro, and what they thought about the timing of the change to the euro and about legislation surrounding the euro. Another set of questions queried respondents about financial services and obstacles to using those services, means of payment for important purchases, bank accounts, post office accounts, loans, and mortgages. Standard demographic information was collected on left-right political self-placement, marital status, age at completion of education, current age, sex, number of people in the household, number of children in the household, current occupation, previous occupation, household income, size of locality, and region of residence.

HOME

SITE MAP

SEARCH

SPECIAL TOPIC  
ARCHIVES

MEMBERSHIP

HELP

# Near-term next steps

- Add records for geospatial data
- Ability to sort or separate results to distinguish GIS and non-spatial data
- Limit search by media type
- Continue to catalog data on the Internet
- Interoperability with other catalogs

# Long-term next steps

- Link study description to live data sets, including documentation and software setups.
- Spatial queries
- Search variables and question text.
- Develop StatCat as a portal to social science numeric data services.