The DataWeb: a new framework for data

Cavan Capps

Chief DataWeb Applications Branch

Data Integration Division

Howard Hogan Director

Demographic Programs Directorate

Data = A number in a context.

“10.0%” is not data

“The 2005 poverty rate for the US is 10.0%” is data

“The 2006 poverty rate for the US as collected by the ACS for the housing unit population is 10.0% (+/- xx%)” is more data

Information on questionnaire, sample size, rotation, imputation, weighting, etc is still more data

The Wider Context

- One Datum is seldom useful
- Analysis requires putting the data point in context.
- Related variables
- Other geographies
- Other time periods

Dissemenation Challenge

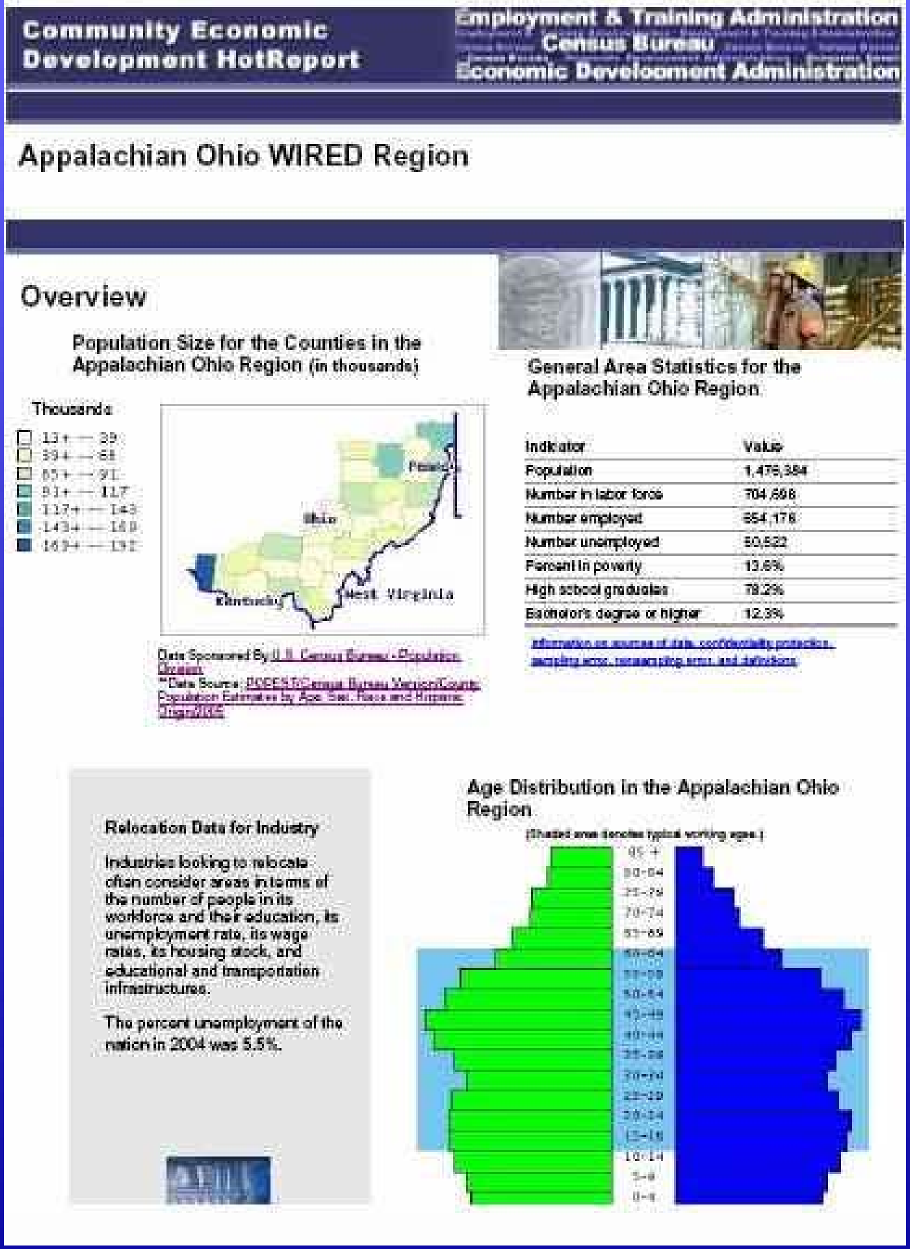
- How to present the right data with the right context to meet users actual needs.
- How to ensure that the most recent and most correct data are displayed

A Three Part Approach

- Hot Reports
- Data Ferrett
- The Data Web

Hot Reports

- Targeted a local decision makers with limited time and statistical background
- Bring together relevant variables for local areas
- Topically oriented
- Updated dynamically
- Can be designed to support decision making
- Guided use of statistical data



Community Economic Development HotReport

Employment & Training Administration
Census Bureau
Economic Development Administration

Appalachian Ohio WIRED Region

Overview

Population Size for the Counties in the Appalachian Ohio Region (in thousands)

Data Sponsored By U.S. Census Bureau - Population Division
**Data Source: POPEST/Census Bureau Version/County Population Estimates by Age, Sex, Race and Hispanic Origin/2005

General Area Statistics for the Appalachian Ohio Region

Indicator	Value
Population	1,476,384
Number in labor force	704,698
Number employed	654,176
Number unemployed	50,522
Percent in poverty	13.6%
High school graduates	78.2%
Bachelor's degree or higher	12.3%

Information on sources of data, confidentiality protection, sampling error, nonsampling error, and definitions

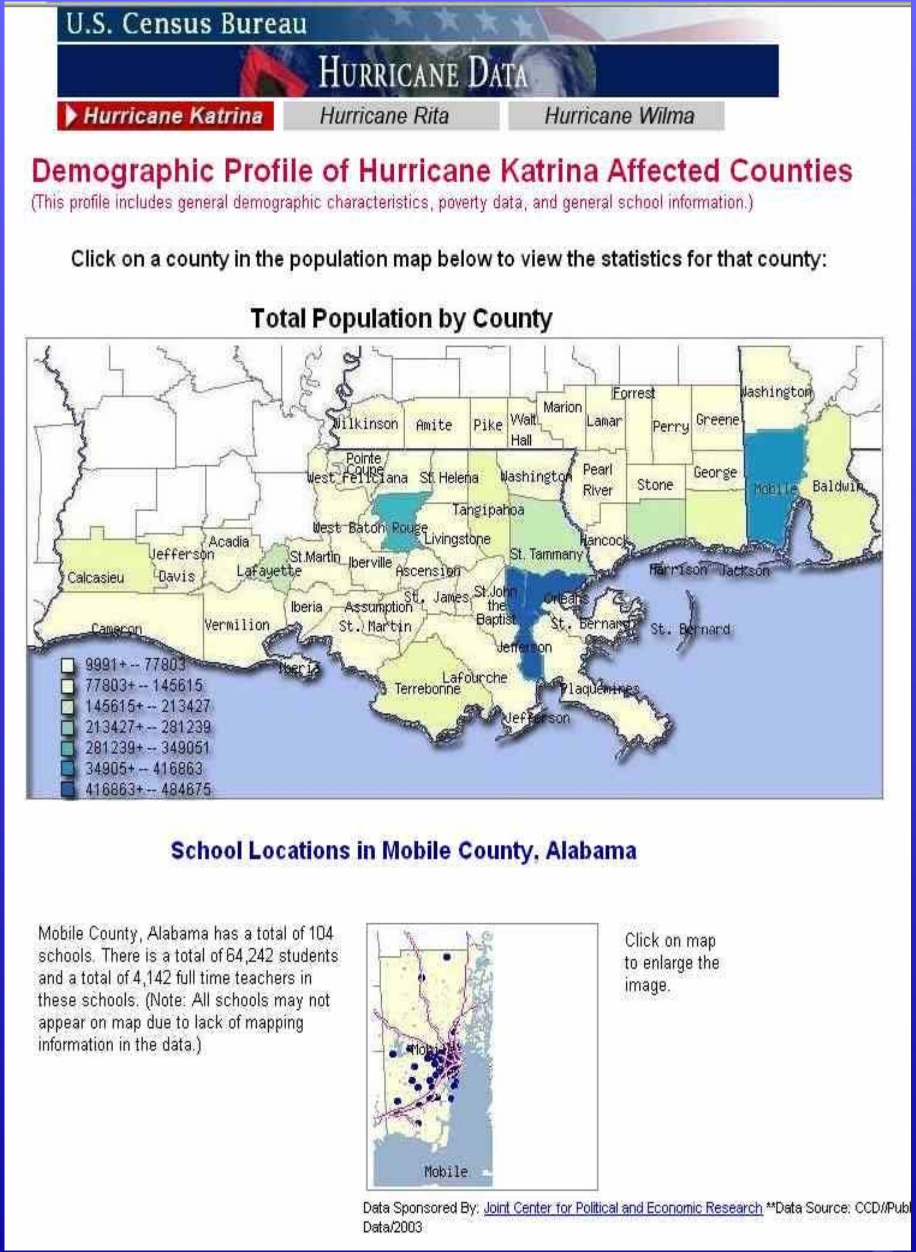
Relocation Data for Industry

Industries looking to relocate often consider areas in terms of the number of people in its workforce and their education, its unemployment rate, its wage rates, its housing stock, and educational and transportation infrastructures.

The percent unemployment of the nation in 2004 was 5.5%.

Age Distribution in the Appalachian Ohio Region

(Shaded area denotes typical working age.)



U.S. Census Bureau

HURRICANE DATA

Demographic Profile of Hurricane Katrina Affected Counties

(This profile includes general demographic characteristics, poverty data, and general school information.)

Click on a county in the population map below to view the statistics for that county:

Total Population by County

School Locations in Mobile County, Alabama

Mobile County, Alabama has a total of 104 schools. There is a total of 64,242 students and a total of 4,142 full time teachers in these schools. (Note: All schools may not appear on map due to lack of mapping information in the data.)

Click on map to enlarge the image.

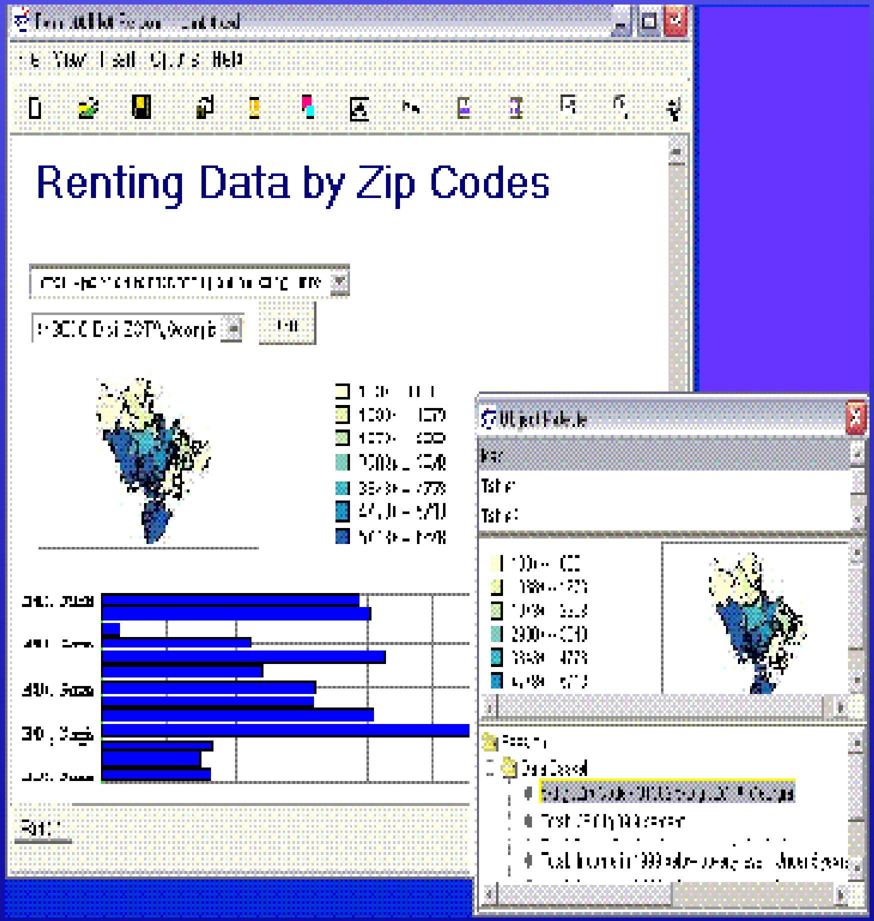
Data Sponsored By: Joint Center for Political and Economic Research **Data Source: CCD/Publ Data/2003

U S C E N S U S B U R E A U

Helping You Make Informed Decisions

Relatively Quick to Build

- Drag & Drop Layout
- Statistically smart
- Gives an analyst a chance to layout data for a problem
- Creates information



Renting Data by Zip Codes

Relatively Quick to Build

- 20% of time is creating Hot Report
- 50% of time is designing Hot Report (finding right data, and laying it out)
- 30% of time is reviewing and fact checking

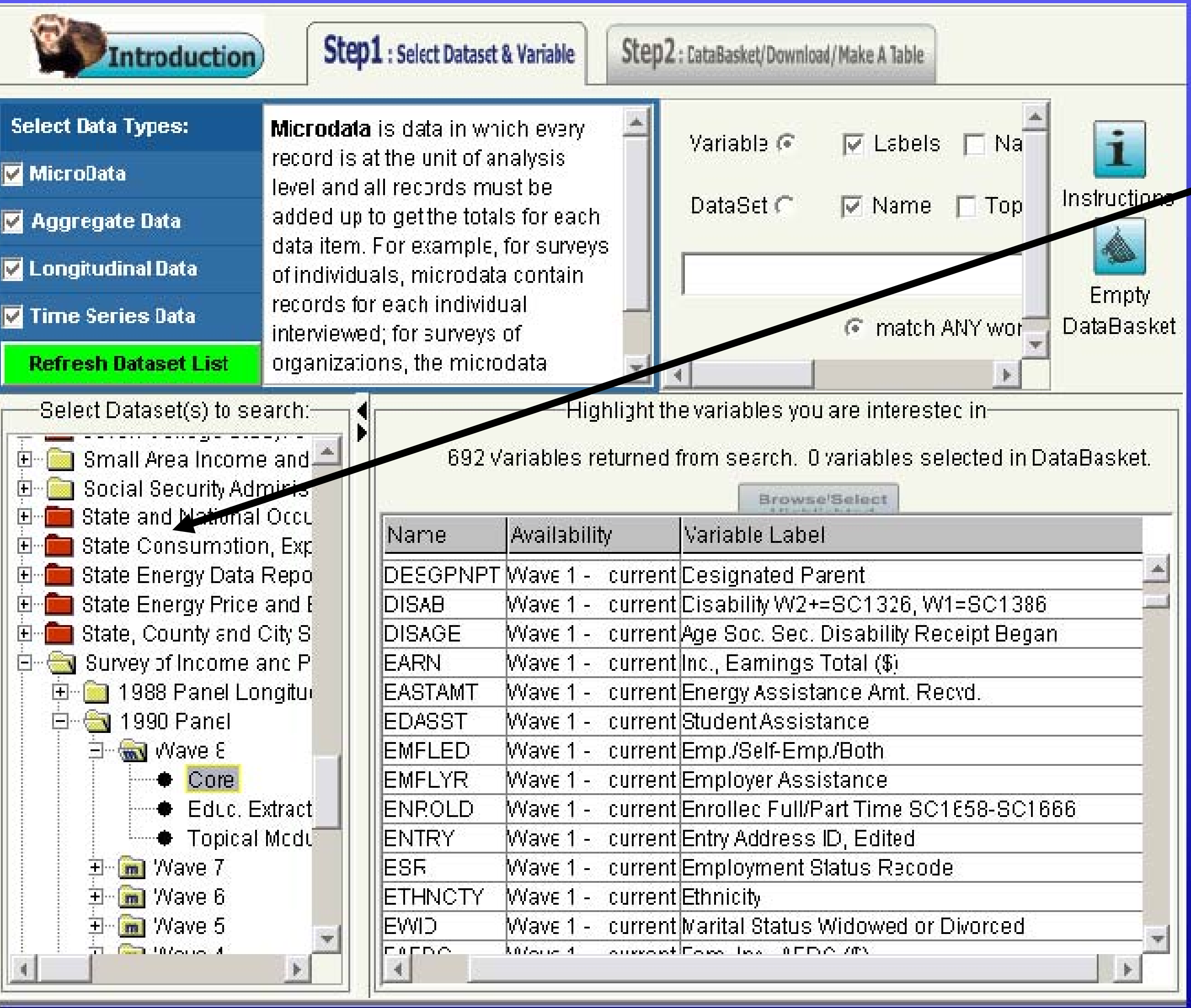
Typical Hot Report Users

- Regional Economic developers
- Emergency Planning and Co-ordination (FEMA)
- Public Health Planning
- Grant Eligibility
- Performance Indicators

Data Ferrett: data browser

- Targeted at sophisticated data users
- Brings together multiple data sets
- Updated dynamically
- Brings data context along with the numbers
- Speeds analysis
  - Data manipulation
  - Advanced Tabulation and descriptive statistics
  - Mapping and business graphics using statistical rules
  - Adding regressions and other advanced statistics

Data Web Browser



Introduction

Step1 : Select Dataset & Variable

Step2 : DataBasket/Download/Make A Table

Select Data Types:

☑ MicroData

☑ Aggregate Data

☑ Longitudinal Data

☑ Time Series Data

Refresh Dataset List

Microdata is data in which every record is at the unit of analysis level and all records must be added up to get the totals for each data item. For example, for surveys of individuals, microdata contain records for each individual interviewed; for surveys of organizations, the microdata

Variable ● ☑ Labels □ Na

DataSet ○ ☑ Name □ Top

● match ANY wor

Instructions

Empty DataBasket

Select Dataset(s) to search:

- Small Area Income and
- Social Security Adminis
- State and National Occu
- State Consumption, Exp
- State Energy Data Repo
- State Energy Price and E
- State, County and City S
- Survey of Income and P
  - 1988 Panel Longitu
  - 1990 Panel
    - Wave 8
      - Core
      - Educ. Extract
      - Topical Modu
    - Wave 7
    - Wave 6
    - Wave 5

Highlight the variables you are interested in

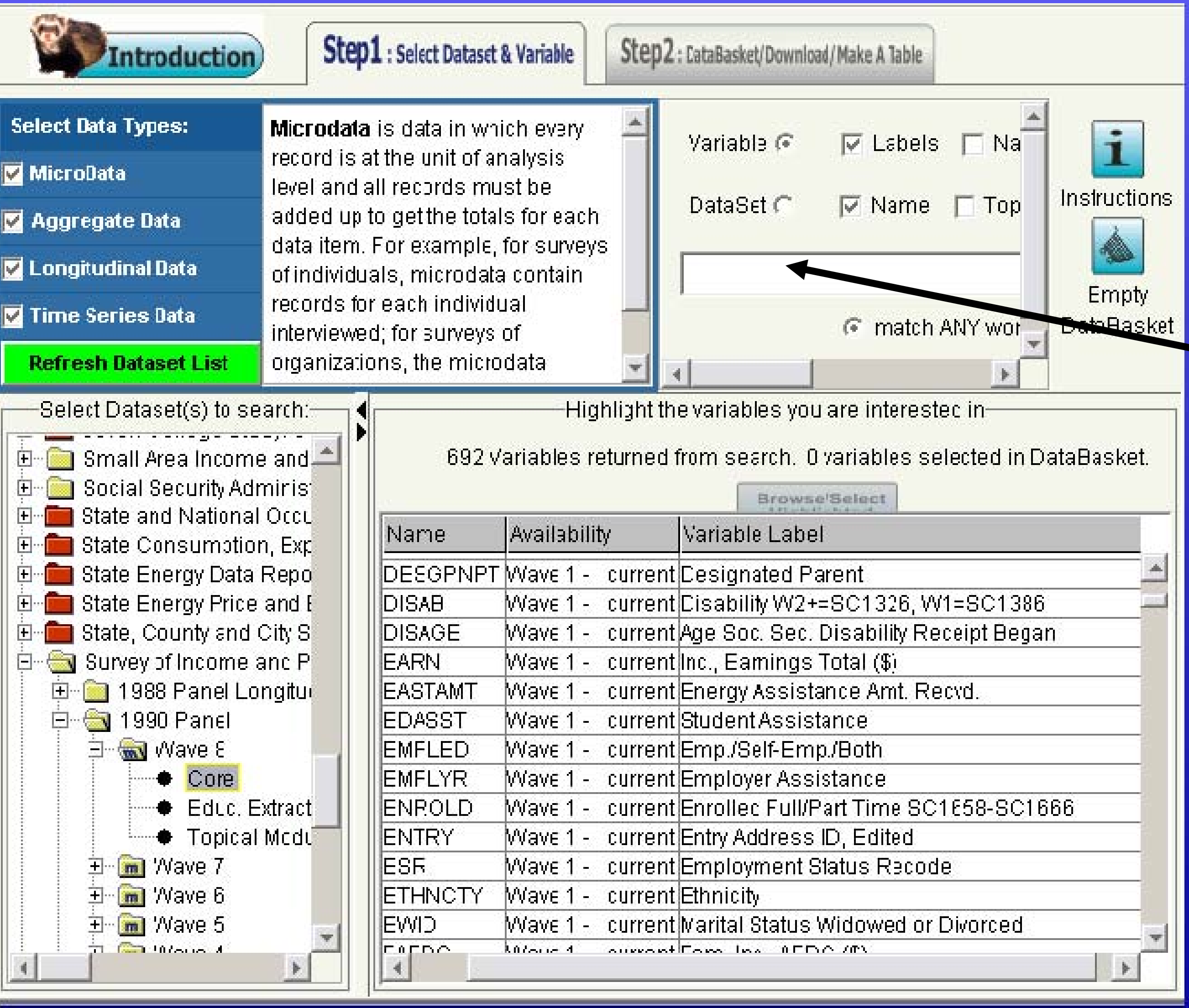
692 Variables returned from search. 0 variables selected in DataBasket.

Browse/Select

Name	Availability	Variable Label
DESGPNPT	Wave 1 - current	Designated Parent
DISAB	Wave 1 - current	Disability W2+=SC1326, W1=SC1386
DISAGE	Wave 1 - current	Age Soc. Sec. Disability Receipt Began
EARN	Wave 1 - current	Inc., Earnings Total ($)
EASTAMT	Wave 1 - current	Energy Assistance Amt. Recvd.
EDASST	Wave 1 - current	Student Assistance
EMFLED	Wave 1 - current	Emp./Self-Emp./Both
EMFLYR	Wave 1 - current	Employer Assistance
ENROLD	Wave 1 - current	Enrolled Full/Part Time SC1658-SC1666
ENTRY	Wave 1 - current	Entry Address ID, Edited
ESR	Wave 1 - current	Employment Status Recode
ETHNCTY	Wave 1 - current	Ethnicity
EWID	Wave 1 - current	Marital Status Widowed or Divorced

- Data set collections are in folders on left

Data Web Browser



Introduction | Step1 : Select Dataset & Variable | Step2 : DataBasket/Download/Make A Table

Select Data Types:

☑ MicroData

☑ Aggregate Data

☑ Longitudinal Data

☑ Time Series Data

Refresh Dataset List

Microdata is data in which every record is at the unit of analysis level and all records must be added up to get the totals for each data item. For example, for surveys of individuals, microdata contain records for each individual interviewed; for surveys of organizations, the microdata

Variable ● ☑ Labels □ Na

DataSet ○ ☑ Name □ Top

● match ANY wor

Instructions

Empty DataBasket

Select Dataset(s) to search:

- Small Area Income and
- Social Security Adminis
- State and National Occu
- State Consumption, Exp
- State Energy Data Repo
- State Energy Price and E
- State, County and City S
- Survey of Income and P
  - 1988 Panel Longitu
  - 1990 Panel
    - Wave 8
      - Core
      - Educ. Extract
      - Topical Modu
    - Wave 7
    - Wave 6
    - Wave 5

Highlight the variables you are interested in

692 Variables returned from search. 0 variables selected in DataBasket.

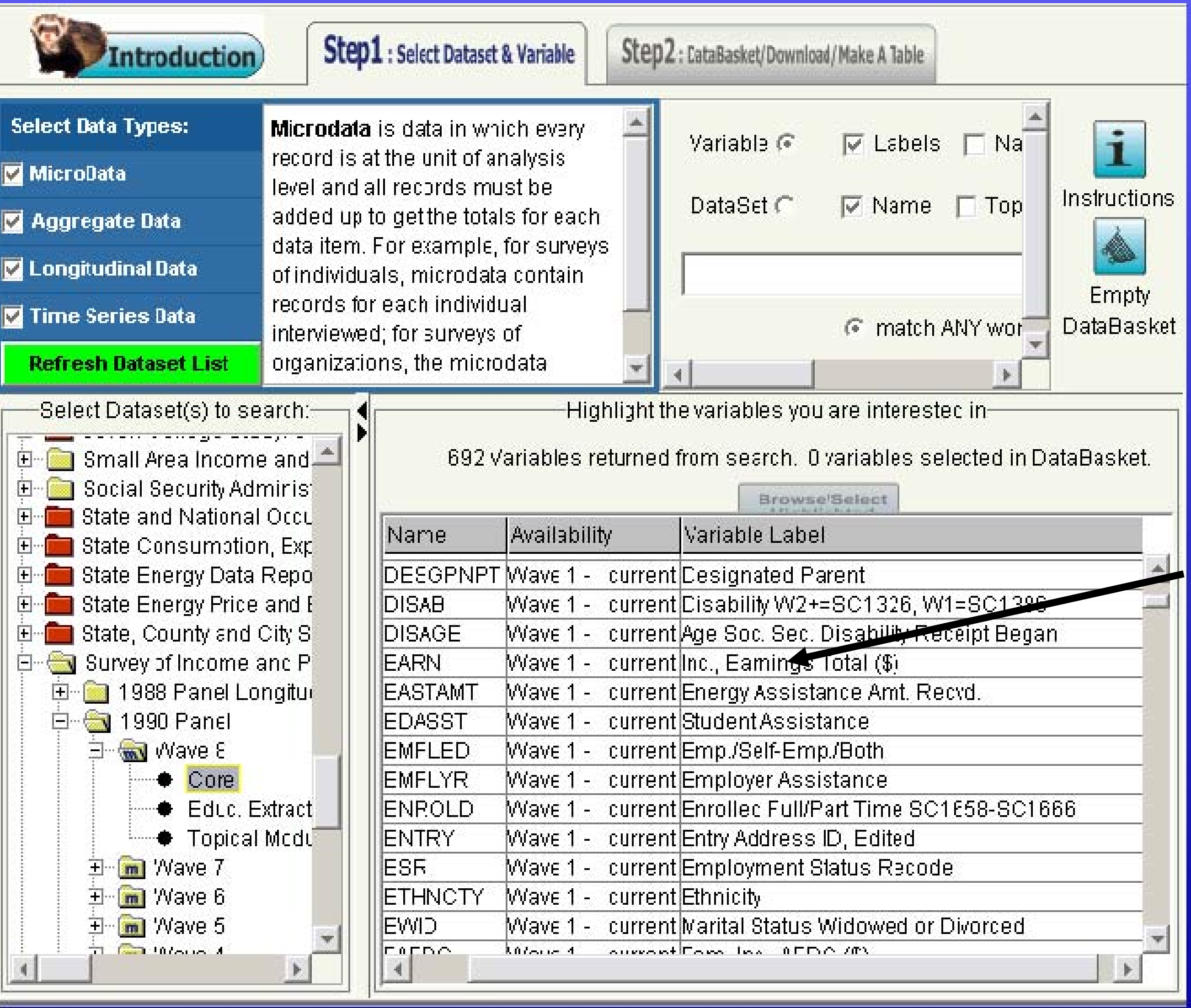
Browse/Select

Name	Availability	Variable Label
DESGPNPT	Wave 1 - current	Designated Parent
DISAB	Wave 1 - current	Disability W2+=SC1326, W1=SC1386
DISAGE	Wave 1 - current	Age Soc. Sec. Disability Receipt Began
EARN	Wave 1 - current	Inc., Earnings Total ($)
EASTAMT	Wave 1 - current	Energy Assistance Amt. Recvd.
EDASST	Wave 1 - current	Student Assistance
EMFLED	Wave 1 - current	Emp./Self-Emp./Both
EMFLYR	Wave 1 - current	Employer Assistance
ENROLD	Wave 1 - current	Enrolled Full/Part Time SC1658-SC1666
ENTRY	Wave 1 - current	Entry Address ID, Edited
ESR	Wave 1 - current	Employment Status Recode
ETHNCTY	Wave 1 - current	Ethnicity
EWID	Wave 1 - current	Marital Status Widowed or Divorced

- Data set collections are in folders on left

- Highlighted Data sets can be searched.

Data Web Browser


Introduction
Step1 : Select Dataset & Variable
Step2 : DataBasket/Download/Make A Table
Instructions
Empty DataBasket
Browse/Select

Select Data Types:

☑ MicroData

☑ Aggregate Data

☑ Longitudinal Data

☑ Time Series Data

Refresh Dataset List

Microdata is data in which every record is at the unit of analysis level and all records must be added up to get the totals for each data item. For example, for surveys of individuals, microdata contain records for each individual interviewed; for surveys of organizations, the microdata

Variable ● ☑ Labels □ Na

DataSet ○ ☑ Name □ Top

● match ANY wor

Select Dataset(s) to search:

- Small Area Income and
- Social Security Adminis
- State and National Occu
- State Consumption, Exp
- State Energy Data Repo
- State Energy Price and E
- State, County and City S
- Survey of Income and P
  - 1988 Panel Longitu
  - 1990 Panel
    - Wave 8
      - Core
      - Educ. Extract
      - Topical Modu
    - Wave 7
    - Wave 6
    - Wave 5

Highlight the variables you are interested in

692 variables returned from search. 0 variables selected in DataBasket.

Name	Availability	Variable Label
DESGPNPT	Wave 1 - current	Designated Parent
DISAB	Wave 1 - current	Disability W2+=SC1326, W1=SC1386
DISAGE	Wave 1 - current	Age Soc. Sec. Disability Receipt Began
EARN	Wave 1 - current	Inc., Earnings Total ($)
EASTAMT	Wave 1 - current	Energy Assistance Amt. Recvd.
EDASST	Wave 1 - current	Student Assistance
EMFLED	Wave 1 - current	Emp./Self-Emp./Both
EMFLYR	Wave 1 - current	Employer Assistance
ENROLD	Wave 1 - current	Enrolled Full/Part Time SC1658-SC1666
ENTRY	Wave 1 - current	Entry Address ID, Edited
ESR	Wave 1 - current	Employment Status Recode
ETHNCTY	Wave 1 - current	Ethnicity
EWID	Wave 1 - current	Marital Status Widowed or Divorced

- Dataset collections are in folders on left

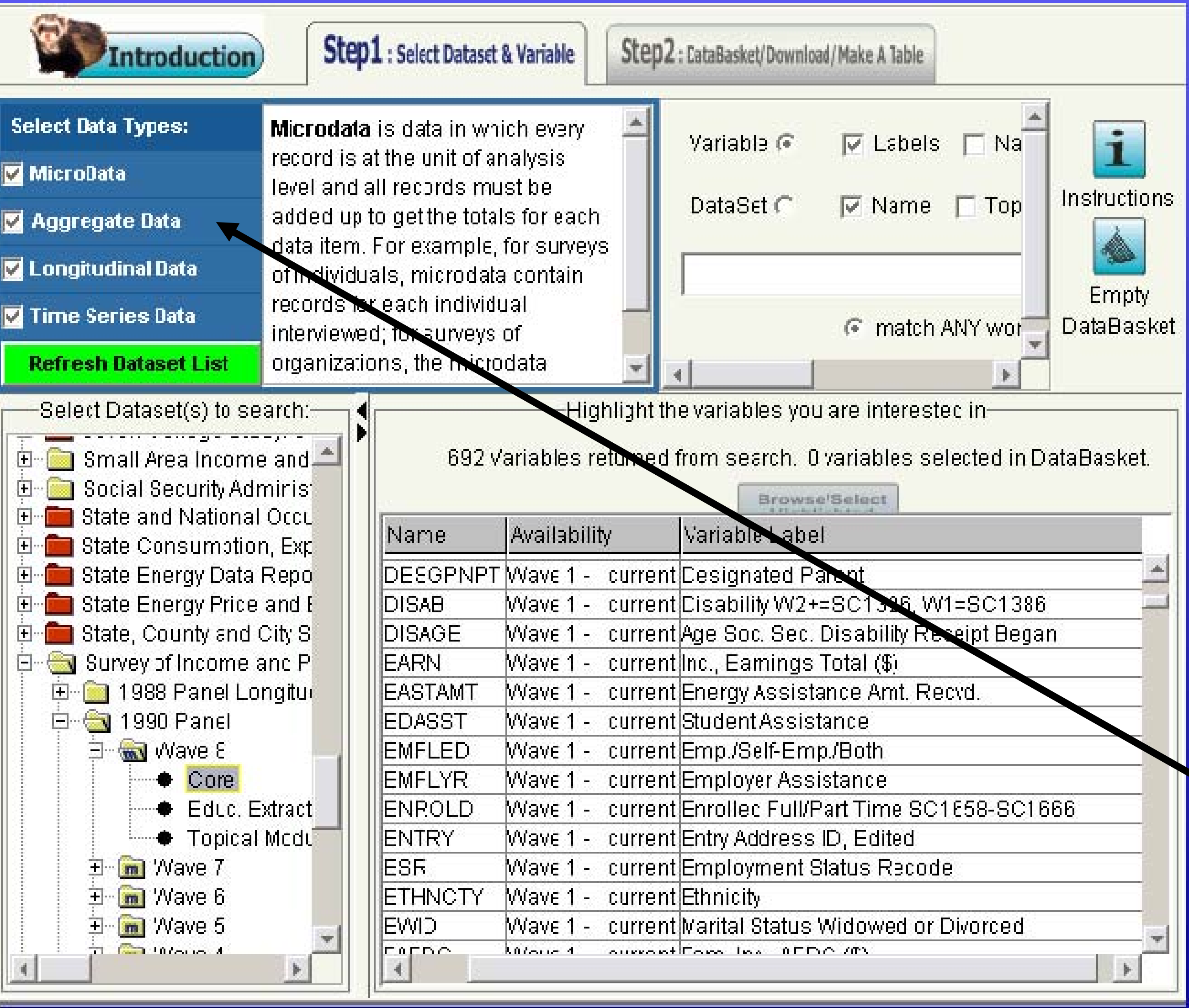
- Highlighted datasets can be searched.

- Variables returned from search

U S C E N S U S B U R E A U

Helping You Make Informed Decisions

Data Web Browser



Microdata is data in which every record is at the unit of analysis level and all records must be added up to get the totals for each data item. For example, for surveys of individuals, microdata contain records for each individual interviewed; for surveys of organizations, the microdata

Select Dataset(s) to search:

- Small Area Income and
- Social Security Adminis
- State and National Occu
- State Consumption, Exp
- State Energy Data Repo
- State Energy Price and E
- State, County and City S
- Survey of Income and P
  - 1988 Panel Longitu
  - 1990 Panel
    - Wave 8
      - Core
      - Educ. Extract
      - Topical Modu
    - Wave 7
    - Wave 6
    - Wave 5

Highlight the variables you are interested in

692 variables returned from search. 0 variables selected in DataBasket.

Name	Availability	Variable Label
DESGPNPT	Wave 1 - current	Designated Parent
DISAB	Wave 1 - current	Disability W2+=SC1386, W1=SC1386
DISAGE	Wave 1 - current	Age Soc. Sec. Disability Receipt Began
EARN	Wave 1 - current	Inc., Earnings Total ($)
EASTAMT	Wave 1 - current	Energy Assistance Amt. Recvd.
EDASST	Wave 1 - current	Student Assistance
EMFLED	Wave 1 - current	Emp./Self-Emp./Both
EMFLYR	Wave 1 - current	Employer Assistance
ENROLD	Wave 1 - current	Enrolled Full/Part Time SC1658-SC1666
ENTRY	Wave 1 - current	Entry Address ID, Edited
ESR	Wave 1 - current	Employment Status Recode
ETHNCTY	Wave 1 - current	Ethnicity
EWID	Wave 1 - current	Marital Status Widowed or Divorced

- Dataset collections are in folders on left.

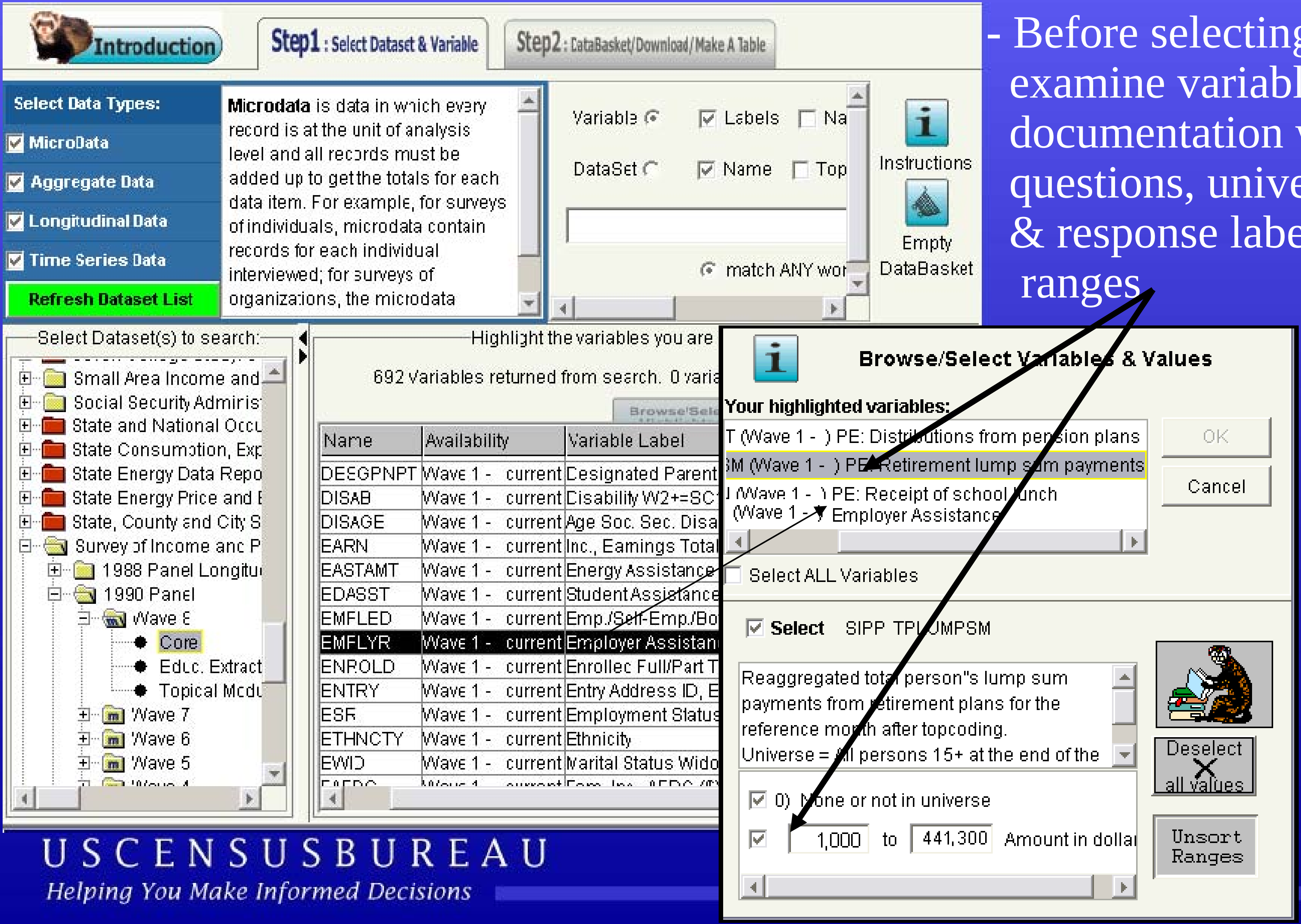
- Highlighted datasets can be searched.

- Variables returned from search.

- Multiple kinds of datasets supported.

Data Web Browser

- Before selecting, examine variable documentation with questions, universes & response labels or ranges


Introduction
Step1 : Select Dataset & Variable
Step2 : DataBasket/Download/Make A Table
Instructions
Empty DataBasket
Browse/Sele

Select Data Types:
☑ MicroData
☑ Aggregate Data
☑ Longitudinal Data
☑ Time Series Data
Refresh Dataset List

Microdata is data in which every record is at the unit of analysis level and all records must be added up to get the totals for each data item. For example, for surveys of individuals, microdata contain records for each individual interviewed; for surveys of organizations, the microdata

Variable ● ☑ Labels □ Na
DataSet ○ ☑ Name □ Top
● match ANY wor

Select Dataset(s) to search:
- Small Area Income and
- Social Security Adminis
- State and National Occu
- State Consumption, Exp
- State Energy Data Repo
- State Energy Price and E
- State, County and City S
- Survey of Income and P
  - 1988 Panel Longitu
  - 1990 Panel
    - Wave 8
      - Core
      - Educ. Extract
      - Topical Modu
    - Wave 7
    - Wave 6
    - Wave 5

Highlight the variables you are

692 Variables returned from search. 0 varia

Name	Availability	Variable Label
DESGPNPT	Wave 1 - current	Designated Parent
DISAB	Wave 1 - current	Disability W2+=SC
DISAGE	Wave 1 - current	Age Soc. Sec. Disa
EARN	Wave 1 - current	Inc., Earnings Total
EASTAMT	Wave 1 - current	Energy Assistance
EDASST	Wave 1 - current	Student Assistance
EMFLED	Wave 1 - current	Emp./Self-Emp./Bo
EMFLYR	Wave 1 - current	Employer Assistanc
ENROLD	Wave 1 - current	Enrolled Full/Part T
ENTRY	Wave 1 - current	Entry Address ID, E
ESR	Wave 1 - current	Employment Status
ETHNCTY	Wave 1 - current	Ethnicity
EWID	Wave 1 - current	Marital Status Wido

Browse/Select Variables & Values

Your highlighted variables:
T (Wave 1 -) PE: Distributions from pension plans
SM (Wave 1 -) PE: Retirement lump sum payments
J (Wave 1 -) PE: Receipt of school lunch
(Wave 1 -) Employer Assistance

OK
Cancel

□ Select ALL Variables

☑ **Select** SIPP TPLUMPSM

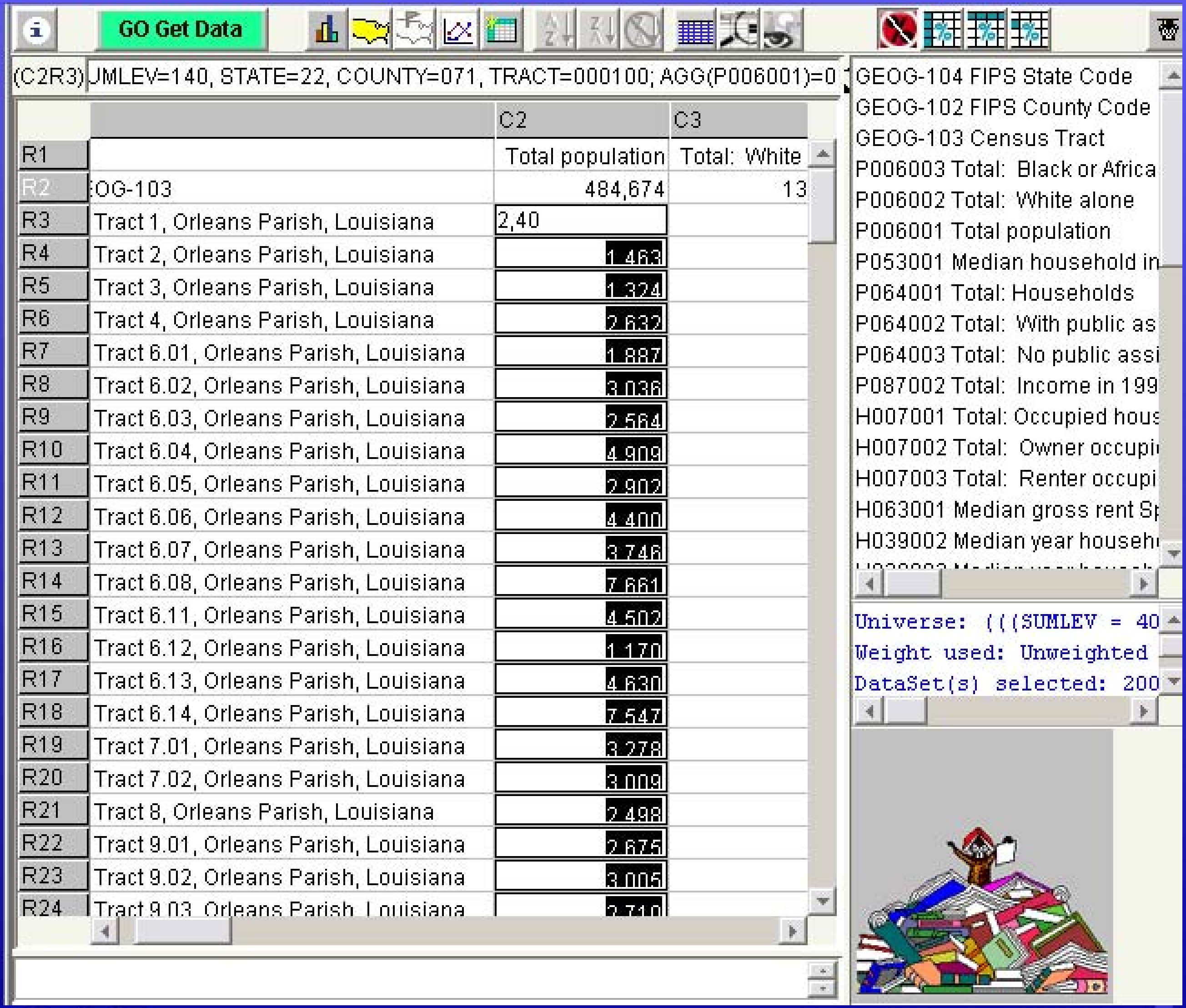
Reaggregated total person"s lump sum payments from retirement plans for the reference month after topcoding.
Universe = All persons 15+ at the end of the

☑ 0) None or not in universe
☑ 1,000 to 441,300 Amount in dollar

Deselect all values
Unsort Ranges

Data Web Browser

- Selected variables are tabulated in the spreadsheet controlled by statistical rules.



(C2R3) JMLEV=140, STATE=22, COUNTY=071, TRACT=000100; AGG(P006001)=0

		C2	C3
R1		Total population	Total: White
R2	:OG-103	484,674	13
R3	Tract 1, Orleans Parish, Louisiana	2,40	
R4	Tract 2, Orleans Parish, Louisiana	1,463	
R5	Tract 3, Orleans Parish, Louisiana	1,324	
R6	Tract 4, Orleans Parish, Louisiana	2,632	
R7	Tract 6.01, Orleans Parish, Louisiana	1,887	
R8	Tract 6.02, Orleans Parish, Louisiana	3,036	
R9	Tract 6.03, Orleans Parish, Louisiana	2,564	
R10	Tract 6.04, Orleans Parish, Louisiana	4,909	
R11	Tract 6.05, Orleans Parish, Louisiana	2,902	
R12	Tract 6.06, Orleans Parish, Louisiana	4,400	
R13	Tract 6.07, Orleans Parish, Louisiana	3,746	
R14	Tract 6.08, Orleans Parish, Louisiana	7,661	
R15	Tract 6.11, Orleans Parish, Louisiana	4,502	
R16	Tract 6.12, Orleans Parish, Louisiana	1,170	
R17	Tract 6.13, Orleans Parish, Louisiana	4,630	
R18	Tract 6.14, Orleans Parish, Louisiana	7,547	
R19	Tract 7.01, Orleans Parish, Louisiana	3,278	
R20	Tract 7.02, Orleans Parish, Louisiana	3,009	
R21	Tract 8, Orleans Parish, Louisiana	2,498	
R22	Tract 9.01, Orleans Parish, Louisiana	2,675	
R23	Tract 9.02, Orleans Parish, Louisiana	3,005	
R24	Tract 9.03, Orleans Parish, Louisiana	2,710	

GEOG-104 FIPS State Code
GEOG-102 FIPS County Code
GEOG-103 Census Tract
P006003 Total: Black or Africa
P006002 Total: White alone
P006001 Total population
P053001 Median household in
P064001 Total: Households
P064002 Total: With public as
P064003 Total: No public assi
P087002 Total: Income in 199
H007001 Total: Occupied hous
H007002 Total: Owner occupi
H007003 Total: Renter occupi
H063001 Median gross rent Sp
H039002 Median year househ

```
Universe: (((SUMLEV = 40
Weight used: Unweighted
DataSet(s) selected: 200
```

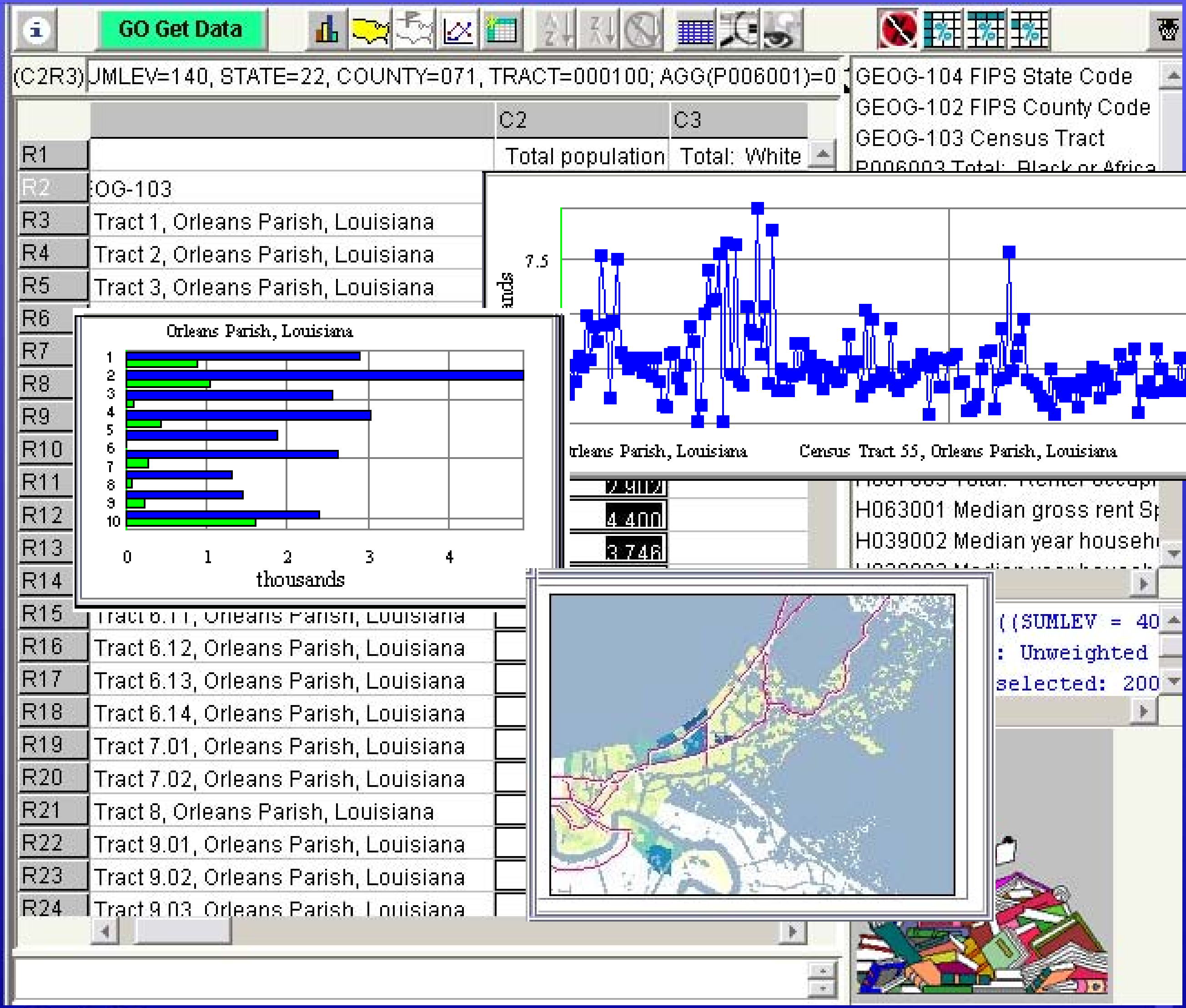
U S C E N S U S B U R E A U

Helping You Make Informed Decisions

Data Web Browser

- Selected variables tabulated in the spreadsheet controlled by the statistical rules.

- Mapping, and business graphics are available for all data.



GO Get Data

(C2R3) JMLEV=140, STATE=22, COUNTY=071, TRACT=000100; AGG(P006001)=0

		C2	C3
R1		Total population	Total: White
R2	OG-103		
R3	Tract 1, Orleans Parish, Louisiana		
R4	Tract 2, Orleans Parish, Louisiana		
R5	Tract 3, Orleans Parish, Louisiana		
R6			
R7			
R8			
R9			
R10			
R11			
R12		2 907	
R13		4 400	
R14		3 746	
R15	Tract 6.11, Orleans Parish, Louisiana		
R16	Tract 6.12, Orleans Parish, Louisiana		
R17	Tract 6.13, Orleans Parish, Louisiana		
R18	Tract 6.14, Orleans Parish, Louisiana		
R19	Tract 7.01, Orleans Parish, Louisiana		
R20	Tract 7.02, Orleans Parish, Louisiana		
R21	Tract 8, Orleans Parish, Louisiana		
R22	Tract 9.01, Orleans Parish, Louisiana		
R23	Tract 9.02, Orleans Parish, Louisiana		
R24	Tract 9.03, Orleans Parish, Louisiana		

GEOG-104 FIPS State Code
GEOG-102 FIPS County Code
GEOG-103 Census Tract
P006003 Total: Black or Africa
H063001 Median gross rent Sp
H039002 Median year househ

```
((SUMLEV = 40
: Unweighted
selected: 200
```

Orleans Parish, Louisiana
1
2
3
4
5
6
7
8
9
10
0
1
2
3
4
thousands
7.5
rleans Parish, Louisiana
Census Tract 55, Orleans Parish, Louisiana

U S C E N S U S B U R E A U
Helping You Make Informed Decisions

Typical Data Ferret Users

- Federal and State government (.gov) - 7876 users and (.us) - 5923 user accounts
- Education (.edu) - 42,828 user accounts
- Non-profit (.org) - 10,792 user user accounts
- Private Companies (.com) – 100,384
  - Press
  - Counsulting (McKinsey, ABT etc.)
  - Retail
  - Marketing
  - Insurance and Financial
  - Pharmaceuticals

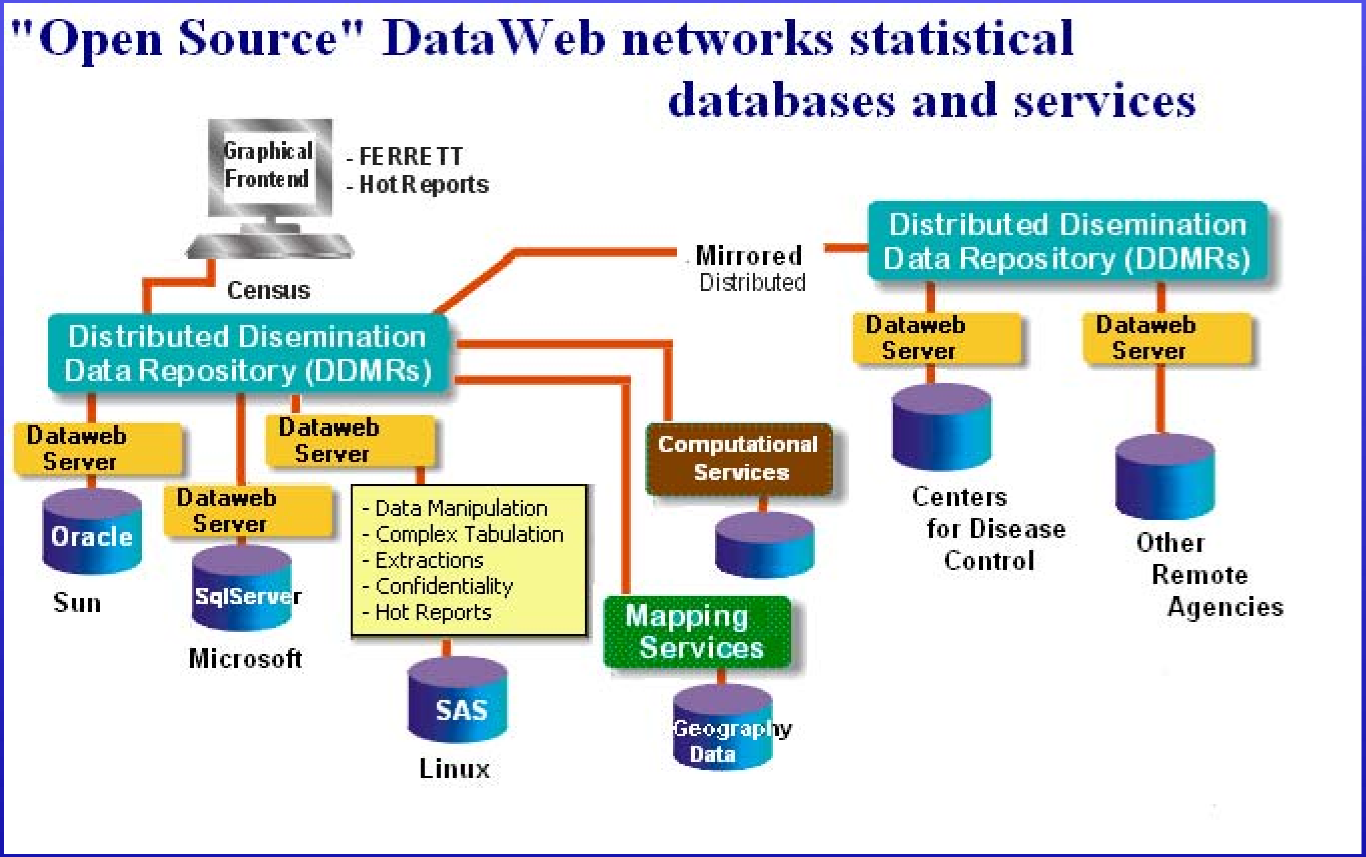
The Data Web

- “The Data Web” is the software engineering that make Data Ferrett and Hot Reports Possible.

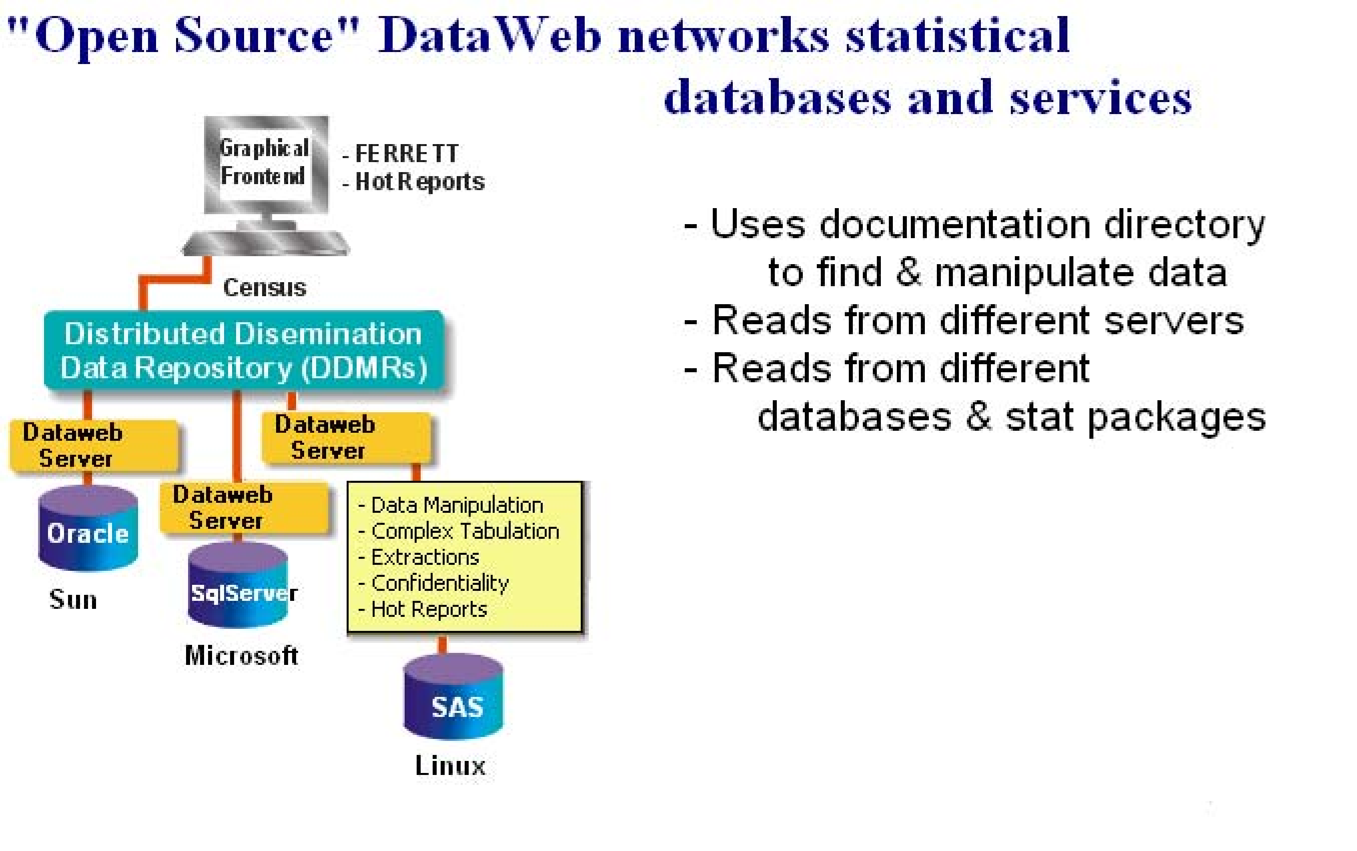
A Smart Data-networking Framework

- Capacity to handle different kinds of data in the same environment or framework
- Empowered by statistical intelligence
  - documentation
  - statistical usage rules
  - data integration rules
- Stores the data one time, use it many times
- More data in the network the more useful

DataWeb Framework



DataWeb Framework

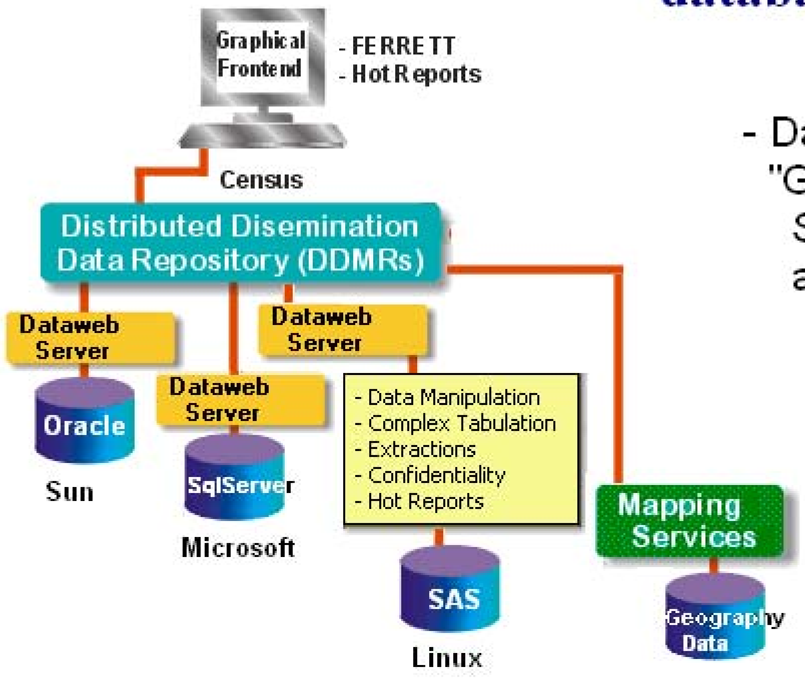


"Open Source" DataWeb networks statistical databases and services

- Uses documentation directory to find & manipulate data
- Reads from different servers
- Reads from different databases & stat packages

DataWeb Framework

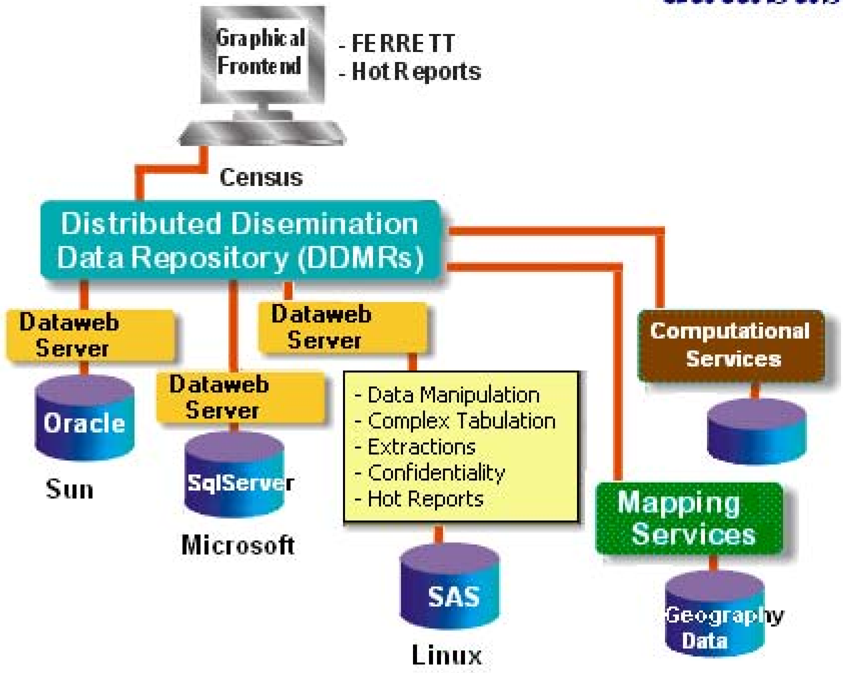
"Open Source" DataWeb networks statistical databases and services



- Data that can be mapped use "Geographical Information Services" (GIS) to create appropriate maps.

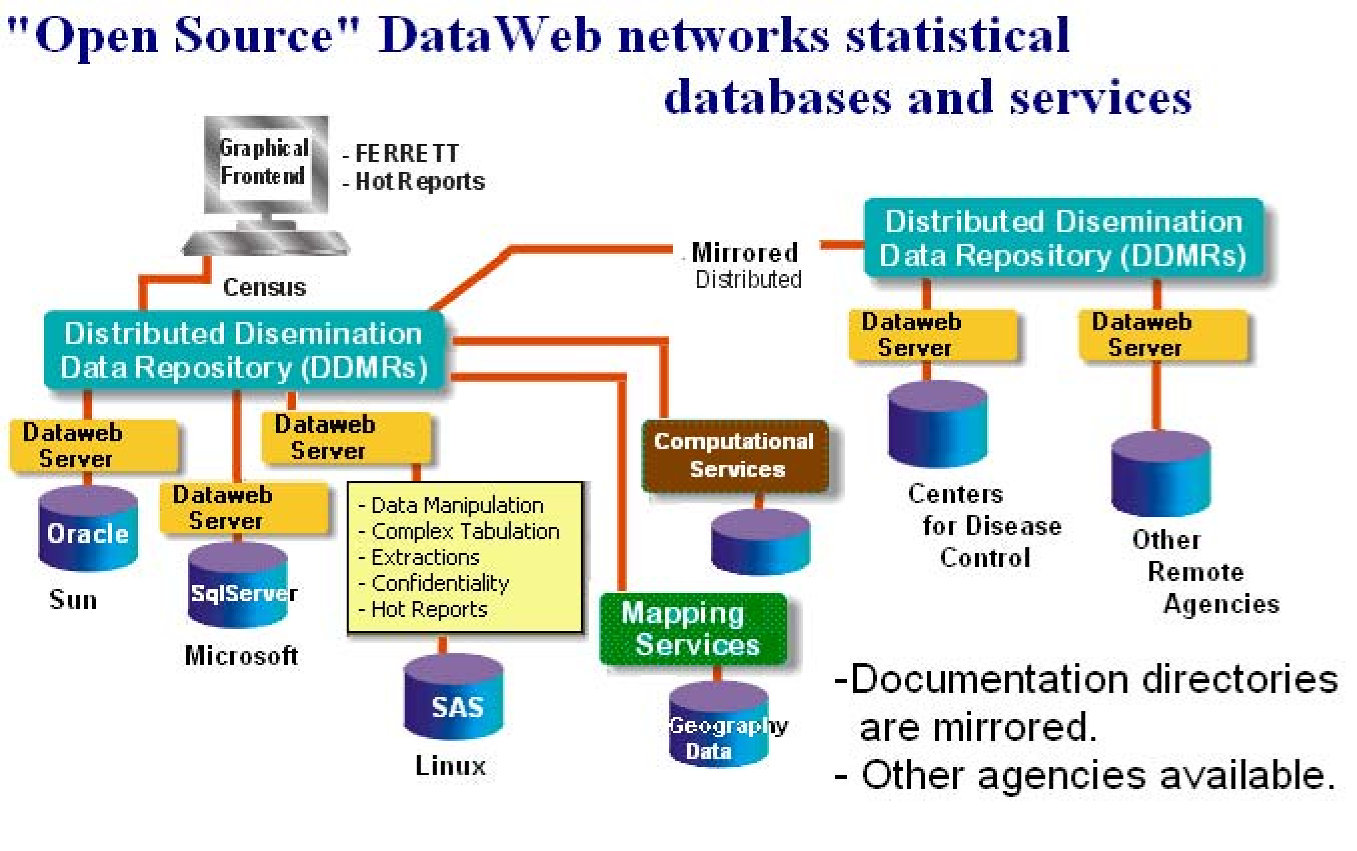
DataWeb Framework

"Open Source" DataWeb networks statistical databases and services



- Computational Services are being added to support advanced statistical modeling for university and advanced users.

DataWeb Framework



"Open Source" DataWeb networks statistical databases and services

-Documentation directories are mirrored.
- Other agencies available.

Based on Collaboration

- “Open Source” statistical partnership with Australian Bureau of Statistics and other interested agencies
- Based on statistical analysts providing statistical rules
- Based on Analysts creating a presentation and analytical review

Important Links

- dataferrett.census.gov
- www.thedataweb.org
- www.thedataweb.org/twiki
- www.thedataweb.org/forum

Cavan Capps

Cavan.Paul.Capps@census.gov