

Toward Digital Government: The Case of Government Statistics

Gary Marchionini
University of North Carolina at Chapel Hill
www.ils.unc.edu/govstat

Williamsburg, VA
September 29, 2005

**South Atlantic Region Conference
Special Libraries Association**

NSF Grants EIA 0131824 and EIA 0129978
Principal Investigators: Gary Marchionini, Stephanie Haas, Ben Shneiderman, Catherine Plaisant, and Carol Hert

Digital Government: 3 Ways to Leverage IT

- Government information dissemination
 - Depository libraries
 - Other publications (no mass emailings yet)
 - QA services
 - [Websites](#) (from tapes to ftp to WWW)
- Transactions
 - Registrations
 - Census, regulatory filings
 - Taxes
- Policy making
 - E-voting
 - E-rules
- Our work focuses on statistical information and agencies as many important decisions by policy makers and citizens depend on statistics

Preliminary Work

1996-2000 (BLS, FedStats)

- Human needs
 - Interviews (agencies, public)
 - Transaction log analysis
 - Email content analysis
- System development and testing
 - Novel interfaces
 - Information architecture
 - Usability studies

Focus on Tables

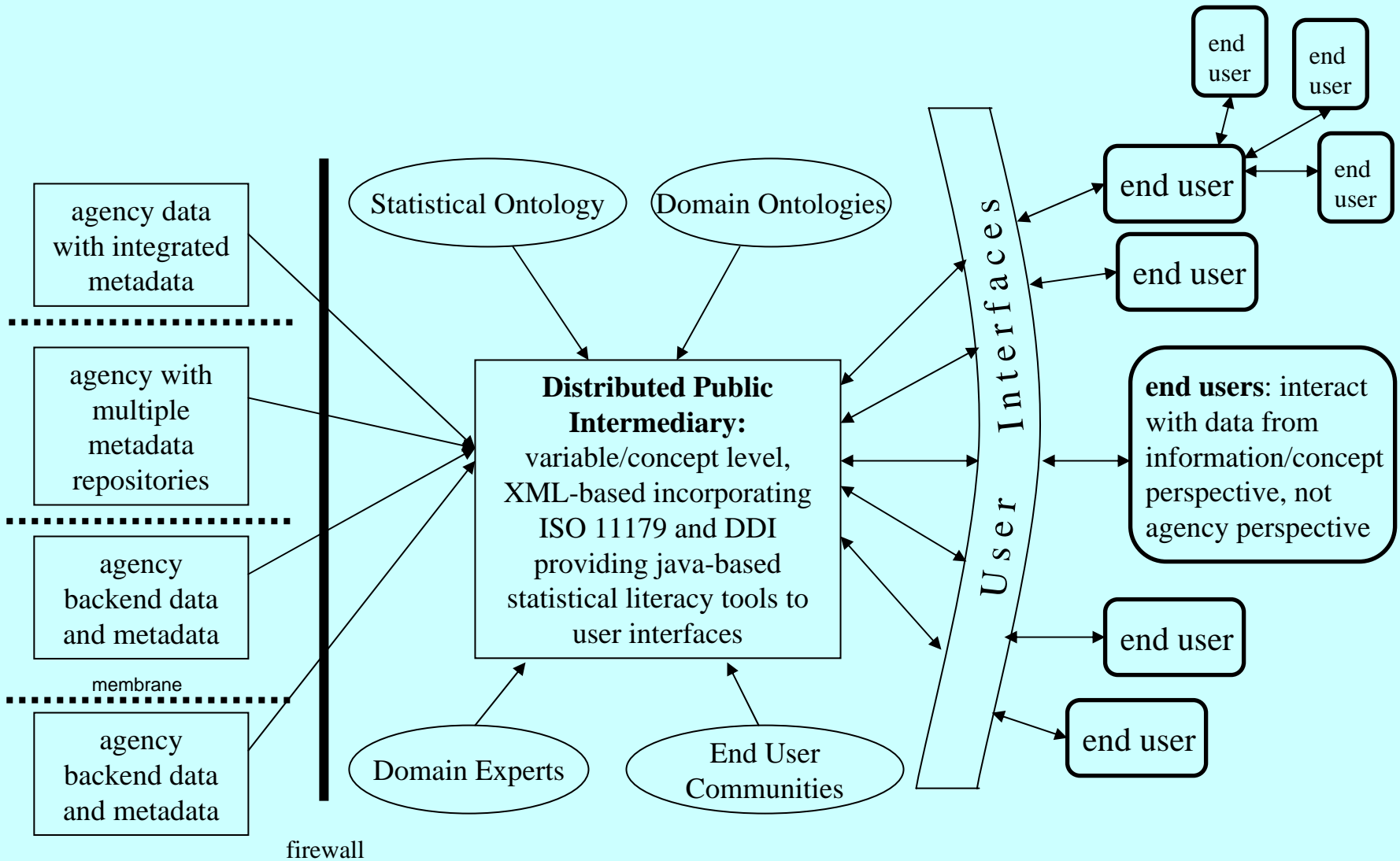
1998-2000 (NSF)

- Table browser
 - Java applet
 - DTD for tables (DC and DDI influence)
 - XML protocol
 - Mapping metadata elements to interface control mechanisms
 - Piping data from large databases to applet
 - User studies
- Metadata to aid understanding

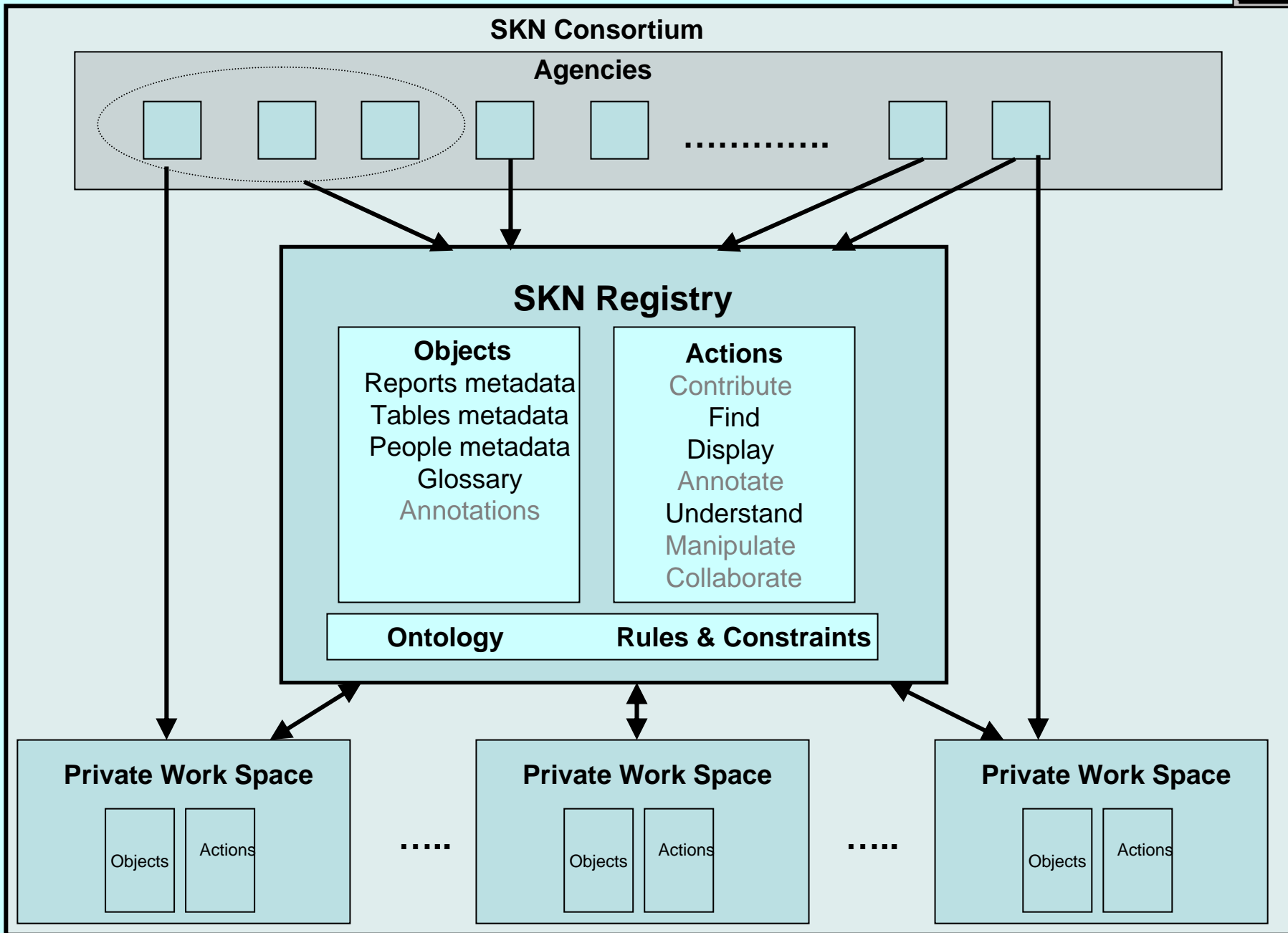
Statistical Knowledge Network 2003-2006 (NSF)

- Create SKN prototype with agency partners
- Integration
 - Horizontal integration across federal agencies (BLS, EIA, NCHS, Census, SSA, NASS)
 - Vertical integration from local/state
- Focus on non-specialists
 - Help crucial
 - Metadata drives help
- User interfaces are the intermediaries to link people and data
- Find what you need, understand what you find

Data Flow



Statistical Knowledge Network Architecture



Interface Prototypes:

*Find, Display, Understand; Leverage
Metadata, Glossary, Ontology*

- Relation Browser
- Multilayered help
 - video help
 - Animated Glossary
- Spatial audio for maps
- Missing Data

Use Case Scenarios to Guide Design

- Based on discussions with agency partners
- 20 scenarios
- 4 detailed with in depth resources located
- Used to ground ongoing work

Relation Browser++ displaying all webpages EIA

EIA Web Collection

Fuel Type: **Alternatives** (126) | **Coal** (905) | **Electricity** (946) | **Natural Gas** (2916) | **Nuclear** (703) | **Petroleum** (834) | **Renewable** (334)

Geography: **State** (1122) | **Region** (729) | **U.S.** (855) | **International** (775)

Sector: **Commercial** (575) | **Electric Utility** (328) | **Industrial** (512) | **Residential** (902)

Process: **Delivery** (541) | **Imports/exports** (403) | **Price/Cost** (942) | **Production** (897) | **Resources/reserves** (703) | **Usage** (774)

2916 result(s) Restart Fewer Categories << More Categories >>

Title	Page Size	URL	Description
Shares of Foreign Direct Investment Position in US Petrole...	4k		... Statistics on Foreign Direct Investment in the United Stat...
International Energy Outlook 2001 - Notes & Sources	26k		... intensive fossil fuel possible, coal, and the least carbon-l...
Gross Withdrawals From Gas and Oil Wells Natural Gas St...	null		... Trillion Cubic Feet) Figure Gross Withdrawals From Gas ...
Figure 6.5 Natural Gas Consumption by Sector	null		... Administration/Annual Energy Review 2001 Figure 6.5 N...
Executive Summary	7k		... billion was spent for natural gas, \$1.4 billion was spent f...
19. Natural Gas Deliveries to Commercial Consumers by S...	null		... Natural Gas Deliveries to Commercial Consumers by St...
Projected Natural Gas Consumption for Electricity Generati...	2k		Slide 16 of 20.
- Natural Gas 1999 NewHampshire New Hampshire - Tabl...	null		... Summary Statistics for Natural Gas New Hampshire, 19...
Highlights Highlights	null		... Natural gas futures prices on the New York Mer- cantile ...
US Natural Gas Plant Processing	12k		... Gas Processed and Liquids Extracted by State, PDF, OT...
15. Consumption of Natural Gas by State, 1993-2000 (Milli...	null		... Consumption of Natural Gas by State, 1993-2000 (Million...
Table E6.2. End Uses of Fuel Consumption, 1998; Level: N...	null		... a Electricity Residual Fuel Oil Distillate Fuel Oil and Dies...
- Natural Gas 1998 NewYork New York -Table 79	null		... Administration / Natural Gas Annual 1998 156 - Natural ...
August Natural Gas Monthly	null		... Average Price of Natural Gas Sold to Industrial Consume...
EIA Environmental Page (Non-Java Version)	30k		... This page links to various US and international legislatio...
Energy Policy Act Transportation Study: Interim Report on N...	null		... 25 5. Percent of End-Use Natural Gas Consumption by ...
The FRS Companies~ Importance in the US Economy	17k		... 2). The bulk of the FRS companiesGÇÖ assets and new ...
- Natural Gas 1999 Tennessee Tennessee -Table 83	null		... Summary Statistics for Natural Gas Tennessee, 1995-19...
Glossary	null		... asphalts. Associated gas: Natural gas found mixed with ...
Weekly Petroleum Status Report	null		... Total commercial petroleum inventories over the last two ...

(Fuel Type=Natural Gas)****

RB++ with Cursor Over Residential Sector



EIA Web Collection

Fuel Type: [v] Geography: [v] Sector: [v] Process: [v]

84	Alternatives	540	State	434	Commercial	193	Delivery
365	Coal	226	Region	177	Electric Utility	205	Imports/exports
404	Electricity	252	U.S.	377	Industrial	407	Price/Cost
902	Natural Gas	189	International	902	Residential	314	Production
315	Nuclear					206	Resources/reserves
280	Petroleum					331	Usage
157	Renewable						

902 result(s) **Restart** Fewer Categories << More Categories >>

Title	Page Size	URL	Description
Energy Glossary, R page	49k It includes the reconfiguration of the vertically-integrated ...
The Northeast Heating Fuel Market: Assessment and Optio...	14k and Pennsylvania. This conforms to Census region 1, co...
- Natural Gas 1996 District of Columbia District of Columbia	null Summary Statistics for Natural Gas District of Columbia, ...
Figure 6.5 Natural Gas Consumption by Sector	null Administration/Annual Energy Review 2001 Figure 6.5 N...
- Natural Gas 1996 Washington Washington	null Vehicle Fuel: Deliveries to Consumers: Electric Resident...
- Natural Gas 1999 NewHampshire New Hampshire -Tabl...	null Summary Statistics for Natural Gas New Hampshire, 19...
- Natural Gas 1998 NewYork New York -Table 79	null Administration / Natural Gas Annual 1998 156 - Natural ...
Ukraine	17k of the country's carbon emissions, with the remainder c...
- Natural Gas 1999 Tennessee Tennessee -Table 83	null Summary Statistics for Natural Gas Tennessee, 1995-19...
Residential Sector Channel	39k Washington, DC 20585. - Home - Petroleum - Gasoline ...
The National Energy Modeling System: An Overview 2000 - ...	13k Energy Information Administration, NEMS International E...
Voluntary Reporting of Greenhouse Gases 1997 - 3. Reduc...	43k or, in the case of CLE Resources, a subsidiary of an elec...
Brunei Country Analysis Brief	26k A second consortium, this one between TotalFinaElf and ...
Retail Unbundling - West Virginia	11k	...	Retail Unbundling - West Virginia. Status: All residential cu...
Energy Plug: Winter Fuels Outlook: 2001-2002	8k Base Forecast. Natural Gas (Midwest). ... Expenditures a...
Energy Savings	11k the savings vary according to the climate in the Census ...
Next Generation * Natural Gas Information Requirements	null For example, the State of Georgia has mandated retail s...
- Natural Gas 1998 Mountain Mountain -Table 42	null Summary Statistics for Natural Gas Mountain, 1994-1998...
Natural Gas Use in American Households	9k Census Region: More than three-fourths of households i...
- Natural Gas 2000 Michigan Michigan -Table 64	null Summary Statistics for Natural Gas Michigan, 1996-2000...

(Fuel Type=Natural Gas)****(Sector=Residential)

RB++ showing 'hous' typed in title field

EIA Web Collection

Fuel Type: **Alternatives** (2), **Coal** (24), **Electricity** (36), **Natural Gas** (59), **Nuclear** (13), **Petroleum** (8), **Renewable** (7)

Geography: **State** (7), **Region** (17), **U.S.** (32), **International** (10)

Sector: **Commercial** (14), **Electric Utility** (11), **Industrial** (19), **Residential** (26)

Process: **Delivery** (15), **Imports/exports** (0), **Price/Cost** (11), **Production** (6), **Resources/reserves** (7), **Usage** (36)

59 result(s) **Restart** Fewer Categories << More Categories >>

Title	Page Size	URL
Greenhouse Gases 1987-1994, Text Boxes	11k	... EIA-s natural gas "venting and flaring" data ar...
Table 1. Dollars Saved per household for a 1-; F Lower	null	... in the Northeast Census Region, 1997 (Continued) RSE ...
Voluntary Reporting of Greenhouse Gases 1997 - 2.	45k	... the renewable energy displaced SCEGÇOs marginal nat...
Emissions of Greenhouse Gases in the United States 1997	21k	US Nitrous Oxide Emissions from Stationary Combustion ...
Table 3. Dollars Saved per household for a 1-; F Lower	null	... Division in the South Census Region, 1997 (Continued) ...
housing Characteristics 1993 -- Total US Data	3k	null
Voluntary Reporting of Greenhouse Gases 2000 - Chapter	31k	... For example, a single DSM program reported by an electr...
Voluntary Reporting of Greenhouse Gases 1999 - 3.	34k	... Two projects involved fuels other than natural gas and el...
Table 2.5 household Energy Consumption by Census	null	(Quadrillion Btu, Except as Noted). Census Region 1. 1978...
Table 2. Dollars Saved per household for a 1-; F Lower	19k	... Midwest Census Region - - -
Table 4. Total Energy Expenditures in US households by	null	... Table 4. Total Energy Expenditures in US Households by...
Voluntary Reporting of Greenhouse Gases 1997 - 1.	37k	... Peabody in the coal mining industry, BP America in the p...
Voluntary Reporting of Greenhouse Gases 2000 - Chapter	34k	... Btu per standard cubic foot to calculate the amount of nat...
Voluntary Reporting of Greenhouse Gases 1997 - 3.	43k	... or, in the case of CLE Resources, a subsidiary of an elec...
Table 1. Total Energy Consumption in US households by	null	... Table 1. Total Energy Consumption in US Households b...
Emissions of Greenhouse Gases in the United States 1997	22k	US Nitrous Oxide Emissions from Stationary Combustion ...
Table 3. Dollars Saved per household for a 1-; F Lower	21k	... South Census Region - - -
Space-Heating energy used by households in the	12k	... Natural Gas and LPG, per Household....Table 5.28. The f...
Figure 7. Electricity and Natural Gas Prices: All US	3k	Figure 7. Electricity and Natural Gas Prices: All US Househ...
Voluntary Reporting of Greenhouse Gases 1997	56k	... emissions if hydropower is used to meet baseload dem...

(Fuel Type=Natural Gas)****(Title=hous)

Animated Demonstration Features

C:\Treemap\Treemap-4.0.2\data\nba-no-hierarchy.tms - Treemap 4.0 Data File loaded at 23:03:07 on 07/09/2003

File Options Help

nba-no-hierarchy.tms3

Atlantic Division
Boston Celtics
Pierce S Walk

Central Division
Atlanta Hawks
He J Terry Wri
Charlotte Hornets
B Da Da Mas West
Miami Heat
Gra Ha Jo Ma
Chicago Bulls
Ar Bran Fi H Mer
New Jersey Nets
J Mar N V W
New York Knicks
C Ho J Ri Spr T
Orlando Magic
Ar McGr M
Philadelphia 76ers
Iverso L S
Washington Wizards
Ham L R
Toronto Raptors
Carter C Da O P W

Midwest Division
Dallas Mavericks
E Finle Ho N Now
Denver Nuggets
La Le Mc P Va
Houston Rockets
Fran Mo T
Minnesota Timberwolves
Br E Gam P Sz
San Antonio Spurs
An Dunc Ro
Utah Jazz
Malo M R S St
Vancouver Grizzlies
Abdu Bib Dic R

Pacific Division
Golden State Warriors
B H J Jamiso S
Los Angeles Clippers
M Od Pi
Los Angeles Lakers
Brya F O'Neal
Phoenix Suns
D Kid Ma Rob Ro
Portland Trail Blazers
P S S St Wal W
Sacramento Kings
C D J Stoj Webb
Seattle SuperSonics
B E Le Pa Payto

Details of selected node

Attribute	Value
(Label)	nba-no-hierarchy.tms3
(Number of nodes)	441
Player	

Main Legend Filters Hierarchy

Label: (Name)

Size: Points

Color: Offensive Rebounds

Maximum depth

[0] 1 2 3

Aggregate: Average

Scale: Linear Square Root Logarithmic

Color Binning

Equally spaced bins

65 390 154 51

0 154 307

Min: 153 Max: 307

Bin Number: 2

Lock the number of bins Prev Next

441 nodes out of 441 nodes (100%) are shown

Data File loaded at 23:03:07 on 07/09/2003

Pacific Division
Golden State Warriors
B H J Jamiso S
Los Angeles Clippers
M Od Pi
Los Angeles Lakers
Brya F O'Neal
Phoenix Suns
D Kid Ma Rob Ro

Details of selected node

Attribute	Value
(Label)	nba-no-hierarchy.tms3
(Number of nodes)	441
Player	
Conference	

Main Legend Filters Hierarchy

Label: (Name)

Size: Points

Color: Offensive Rebounds

Maximum depth

[0] 1 2 3

Aggregate: Average

Scale: Linear Square Root Logarithmic

Data File loaded at 23:03:07 on 07/09/2003

Pacific Division
Golden State Warriors
B H J Jamiso S
Los Angeles Clippers
M Od Pi
Los Angeles Lakers
Brya F O'Neal
Phoenix Suns
D Kid Ma Rob Ro

Details of selected node

Attribute	Value
(Label)	nba-no-hierarchy.tms3
(Number of nodes)	441
Player	

Main Legend Filters Hierarchy

Label: (Name)

Size: Points

Color: Offensive Rebounds

Maximum depth

[0] 1 2 3

Aggregate: Average

Scale: Linear Square Root Logarithmic

How did we create
this Treemap ?

Choosing a layout
algorithm

Interactive Glossary Development Tools



- Provide foundation for content development
- Separate content development from presentation development
- Reduce overall development time
- Maximize reuse of existing elements
- Create multiple presentations from a single content development effort

Animation Template

Content Foundation Title

Content Foundation Question

Content Foundation Presentation Area

Content Foundation Text (Answer & Explanation)

Animation Template Navigation

Specific: Consumer Price Index

What is the Consumer Price Index?

The Consumer Price Index (CPI) represents changes in prices of all goods and services produced for consumption by urban households.

Items are classified into 8 major groups: Food and Beverages, Housing, Apparel, Transportation, Medical Care, Recreation, Education and Communication, Other.

The Consumer Price Index has increased

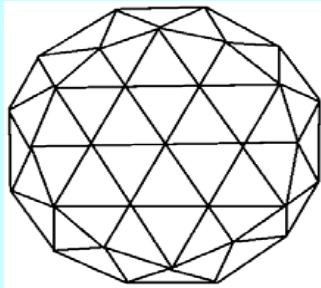
Year	CPI
1995	Low
1996	Low-Mid
1997	Mid
1998	Mid-High
1999	High
2000	High-Mid
2001	Mid-High
2002	High

back play again next

Animation Template

- Three pieces: text, animations, template
- Text is tagged with content section tags in a separate text file
- Animation consists of segments in individual animation files
- Text and animation segments coordinated by placement in template

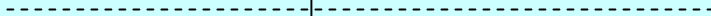
ontology



Semantic level

- Classes
- Relationships
- Constraint rules

modeling



DTD/XML Schema

Structural level

- Elements
- Attributes
- Datatypes

implementation



SKN

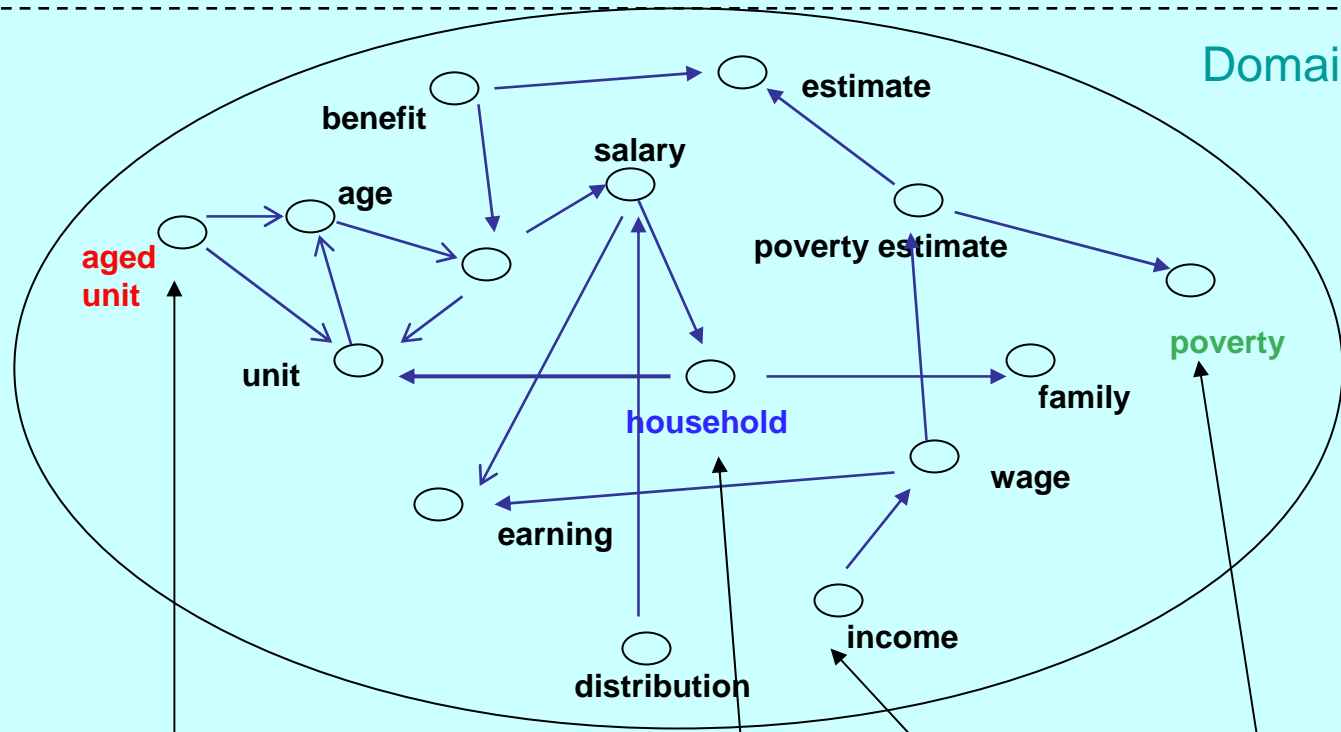
- **Ontology**
- **DTD / XML Schema**
- **Interface Tools**
- **Statistical Interactive Glossary (SIG)**

Ontology Applications

- Knowledge organization
- Content and terminology control
- Data integration
- Query support
- Automatic classification support
- Reasoning mechanism
- Others



Domain knowledge



Operational knowledge

```
<anlyUnit>aged unit</anlyUnit>
<universe>married couples living together,
with husband or wife aged 65 or older</universe>
```



SSA



Census Bureau



FIFARS

Project DTD

- Investigate DDI and ISO 11179
- Leverage DDI and data cubes
- Markup a set of objects
 - Tables
 - Reports/press releases
- Use markup to build added value search (find what you need) and help (understand what you find) support into interfaces

Discovering Metadata

- Hybrid machine learning approach
 - Crawl website
 - Create term document matrices
 - Use k-means clustering with small K to fit on screen in RB++
 - Revise
 - Use structure in the existing sites to train a classifier
 - For small n of concepts, classify site

Combining Machine Learning and Dynamic Interfaces

What should these topics be, and how do we know if we've found the right names for them?

95 result(s) Restart Fewer Categories << More Categories >>

url:

url

http://squash.ils.unc.edu/~cluster/bls_mirror/k12/html/sci_005.htm
http://squash.ils.unc.edu/~cluster/bls_mirror/k12/text/sci_006t.htm
http://squash.ils.unc.edu/~cluster/bls_mirror/k12/text/soc_002t.htm
http://squash.ils.unc.edu/~cluster/bls_mirror/k12/text/sci_005t.htm
http://squash.ils.unc.edu/~cluster/bls_mirror/k12/text/gym_006t.htm
http://squash.ils.unc.edu/~cluster/bls_mirror/k12/text/mat_006t.htm
http://squash.ils.unc.edu/~cluster/bls_mirror/k12/html/sci_005.htm
http://squash.ils.unc.edu/~cluster/bls_mirror/k12/text/gym_006t.htm
http://squash.ils.unc.edu/~cluster/bls_mirror/k12/text/sci_006t.htm
http://squash.ils.unc.edu/~cluster/bls_mirror/k12/text/mat_006t.htm
http://squash.ils.unc.edu/~cluster/bls_mirror/k12/text/soc_002t.htm
http://squash.ils.unc.edu/~cluster/bls_mirror/k12/text/sci_005t.htm

(Main Topics=Occupation) AND (Secondary Topics=Safety)

Java Applet Window

Combining Machine Learning and Dynamic Interfaces

How do we assign thousands of documents to their respective topics?

IDL BLS

Main Topics: Occupation
Secondary Topics: Safety

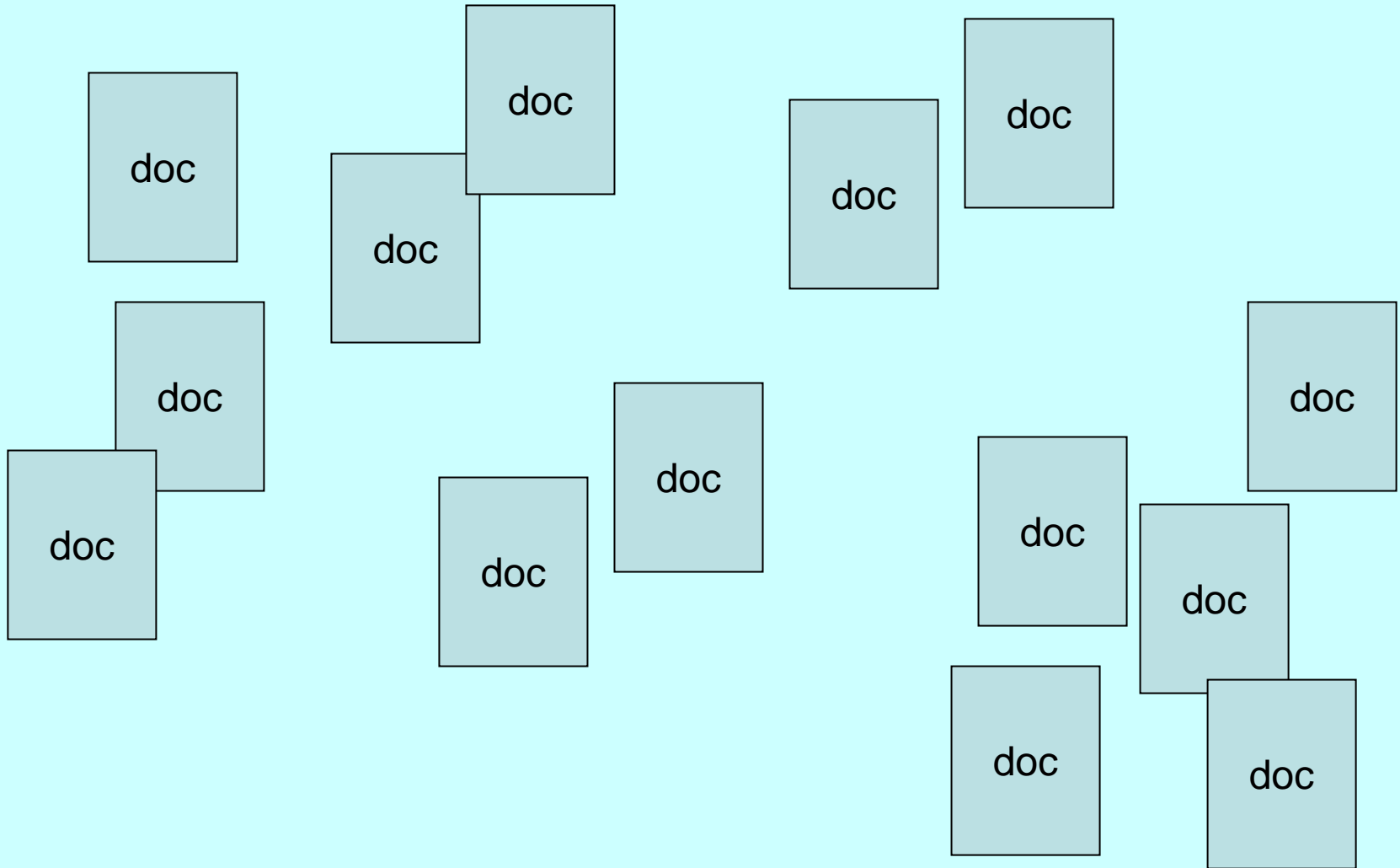
95 result(s) Restart Fewer Categories << More Categories >>

url:
http://squash.ils.unc.edu/~cluster/bls_mirror/k12/html/sci_005.htm
http://squash.ils.unc.edu/~cluster/bls_mirror/k12/text/sci_006t.htm
http://squash.ils.unc.edu/~cluster/bls_mirror/k12/text/soc_002t.htm
http://squash.ils.unc.edu/~cluster/bls_mirror/k12/text/sci_005t.htm
http://squash.ils.unc.edu/~cluster/bls_mirror/k12/text/gym_006t.htm
http://squash.ils.unc.edu/~cluster/bls_mirror/k12/text/mat_006t.htm
http://squash.ils.unc.edu/~cluster/bls_mirror/k12/html/sci_005.htm
http://squash.ils.unc.edu/~cluster/bls_mirror/k12/text/gym_006t.htm
http://squash.ils.unc.edu/~cluster/bls_mirror/k12/text/sci_006t.htm
http://squash.ils.unc.edu/~cluster/bls_mirror/k12/text/mat_006t.htm
http://squash.ils.unc.edu/~cluster/bls_mirror/k12/text/soc_002t.htm
http://squash.ils.unc.edu/~cluster/bls_mirror/k12/text/sci_005t.htm

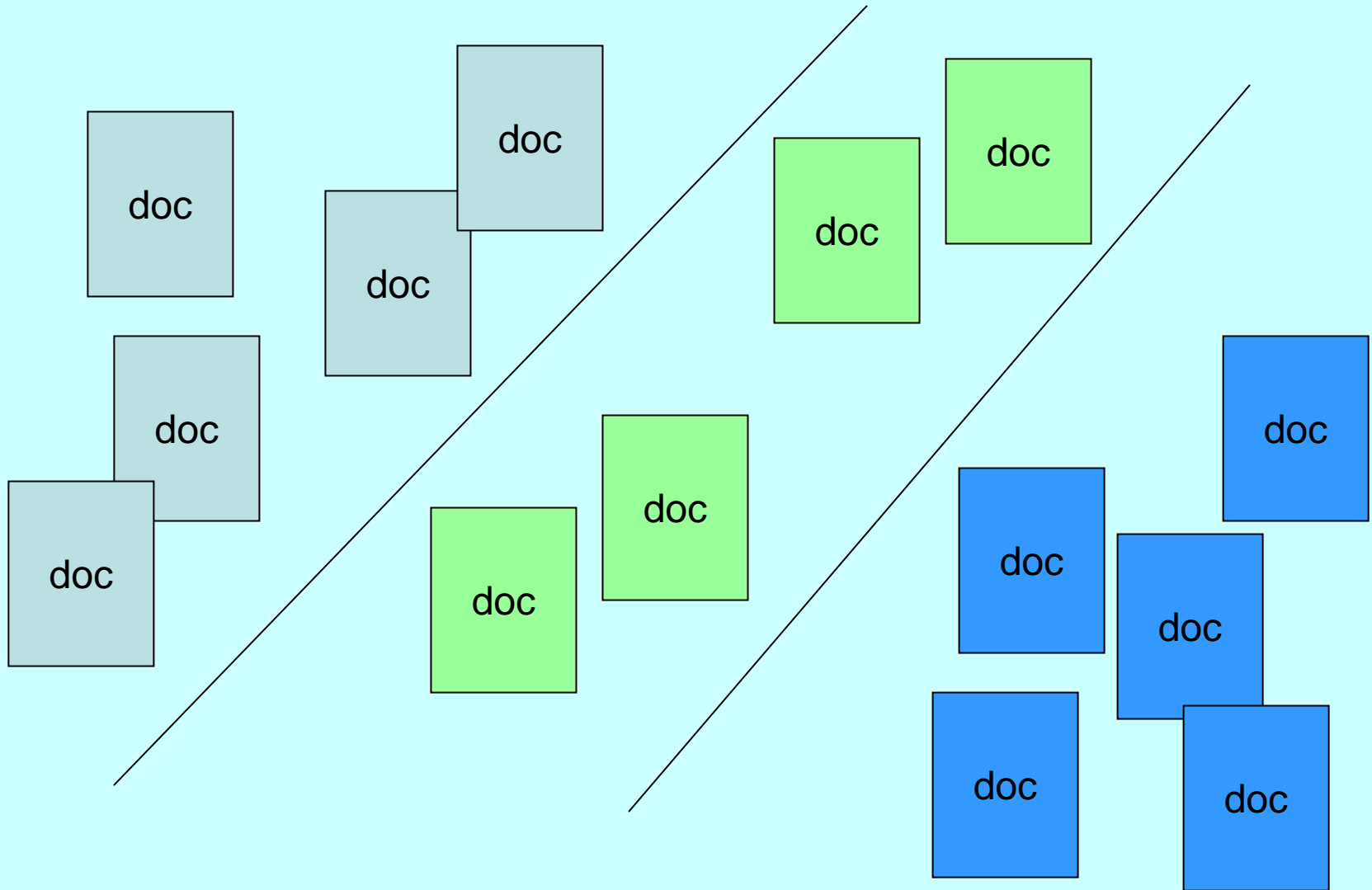
(Main Topics=Occupation) AND (Secondary Topics=Safety)

Java Applet Window

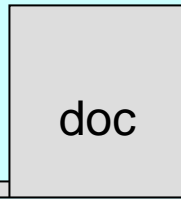
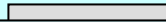
Unstructured Approach



Unstructured Approach

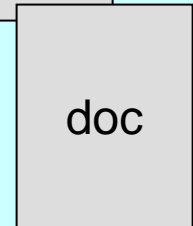


Unstructured Approach



Clustering Based on Word Distributions

This approach yielded intuitively coherent clusters. But the clusters fall at too fine a level of granularity, while also wasting large portions of the data.



Semi-Supervised Approach

Bureau of Labor Statistics Home Page - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address http://www.bls.gov

U.S. Department of Labor
Bureau of Labor Statistics

www.bls.gov Search A-Z Index

About BLS | Jobs in BLS | Get Detailed Statistics | Economic News Releases | Glossary | What's New | Contact Us | Find It! In DOL

Inflation & Consumer Spending
[Consumer Price Index](#) • [Inflation Calculator](#) • [Contract Escalation](#) • [Producer Price Indexes](#) • [Import/Export Price Indexes](#) • [Consumer Expenditures](#) • [Price Index Research](#)

Wages, Earnings, & Benefits
[Wages by Area and Occupation](#) • [Earnings by Industry](#) • [Employee Benefits](#) • [Employment Costs](#) • [State and County Wages](#) • [National Compensation Data](#) • [Collective Bargaining](#)

Productivity
[Productivity and Costs](#) • [Multifactor Productivity](#) • [International Comparisons](#)

Safety & Health
[Injuries and Illnesses](#) • [Fatalities](#)

International
[Import/Export Price Indexes](#) • [Foreign Labor Statistics](#) • [International Technical Cooperation](#)

Occupations
[Occupational Outlook Handbook](#) • [Occupational Outlook Quarterly](#) • [Employment](#) • [Wages by Area and](#)

Latest Numbers

CPI:
unchanged in Oct 2003

Unemployment Rate:
6.0% in Oct 2003

Payroll Employment:
+126,000(p) in Oct 2003

Average Hourly Earnings:
+\$0.01(p) in Oct 2003

PPI:
+0.8%(p) in Oct 2003

ECI:
+1.0% in 3rd Qtr of 2003

Productivity:
+8.1% in 3rd Qtr of 2003

U.S. Import Price Index:
+0.1% in Oct 2003

 » p- preliminary
 » [Subscribe to BLS News](#)
 » [Publication Schedule](#)

Employment & Unemployment
[National Employment](#) • [National Unemployment Rate](#) • [State and Local Employment](#) • [State and Local Unemployment Rates](#) • [Mass Layoffs](#) • [Employment Projections](#) • [Job Openings and Labor Turnover](#) • [Employment by Occupation](#) • [Longitudinal Studies](#) • [State and County Employment](#) • [Time Use](#) • [Business Employment Dynamics](#)

At a Glance Tables
[U.S. Economy at a Glance](#) • [Regions, States, and Areas at a Glance](#) • [Industries at a Glance](#)

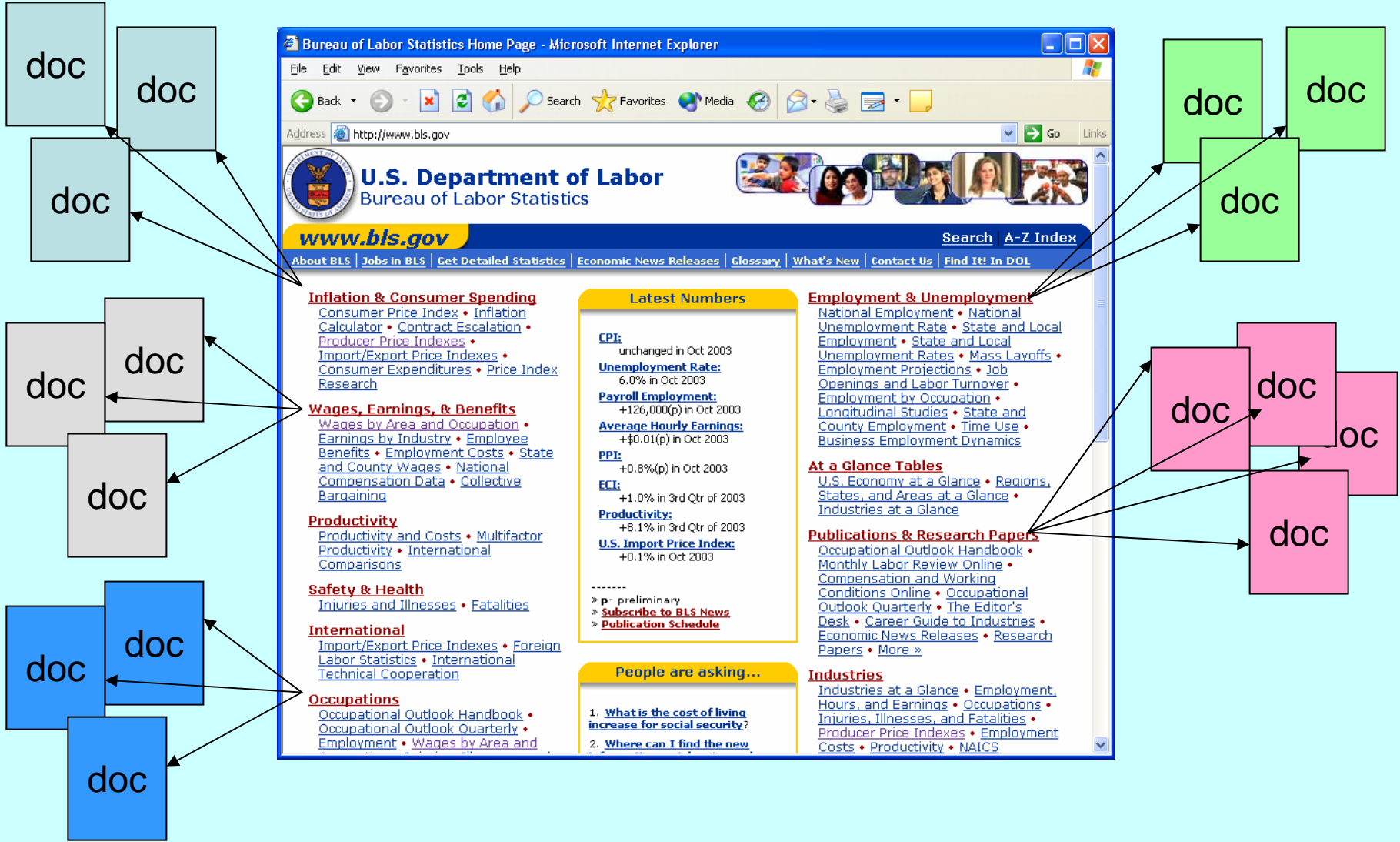
Publications & Research Papers
[Occupational Outlook Handbook](#) • [Monthly Labor Review Online](#) • [Compensation and Working Conditions Online](#) • [Occupational Outlook Quarterly](#) • [The Editor's Desk](#) • [Career Guide to Industries](#) • [Economic News Releases](#) • [Research Papers](#) • [More »](#)

Industries
[Industries at a Glance](#) • [Employment, Hours, and Earnings](#) • [Occupations](#) • [Injuries, Illnesses, and Fatalities](#) • [Producer Price Indexes](#) • [Employment Costs](#) • [Productivity](#) • [NAICS](#)

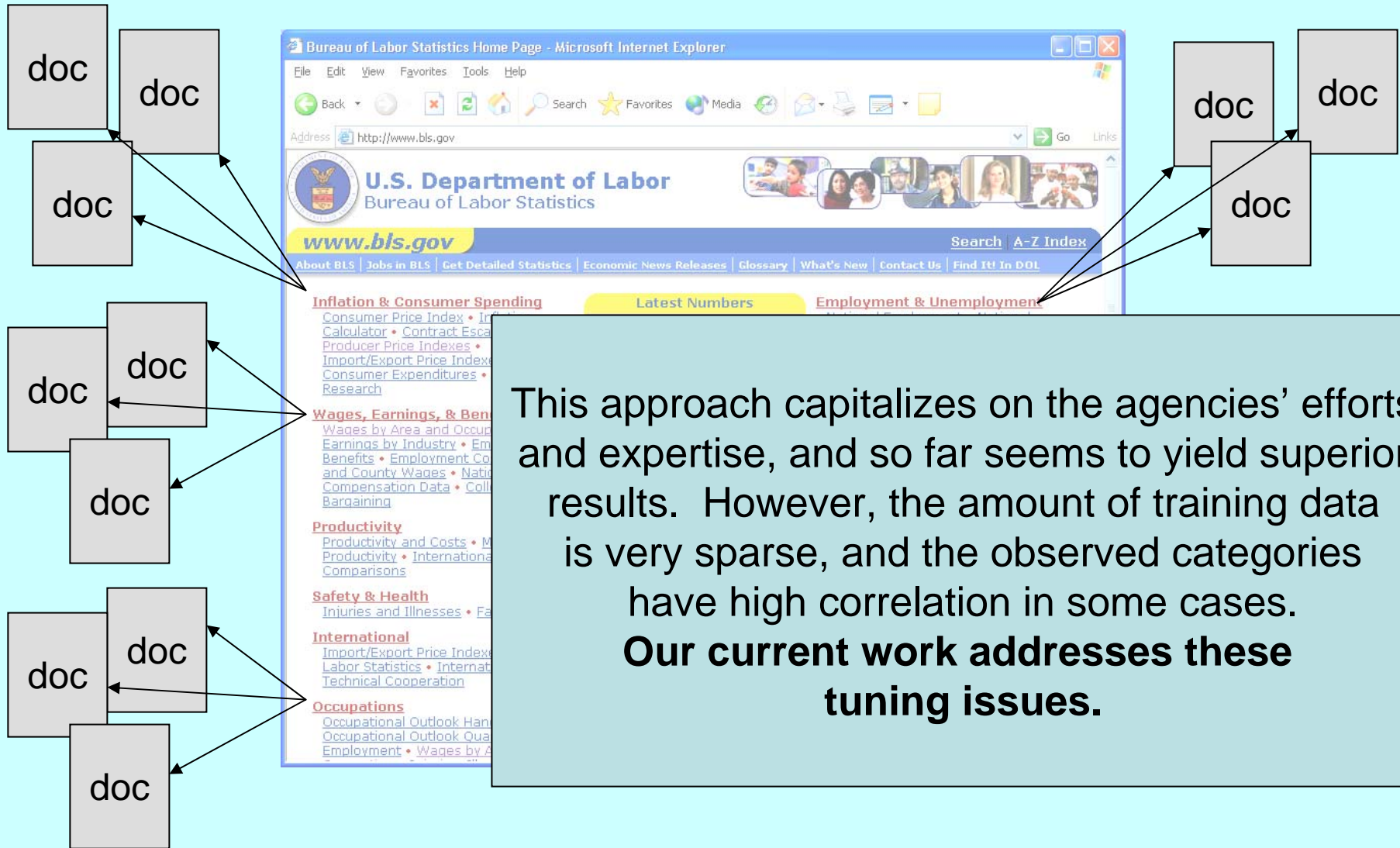
People are asking...

- [What is the cost of living increase for social security?](#)
- [Where can I find the new](#)

Semi-Supervised Approach

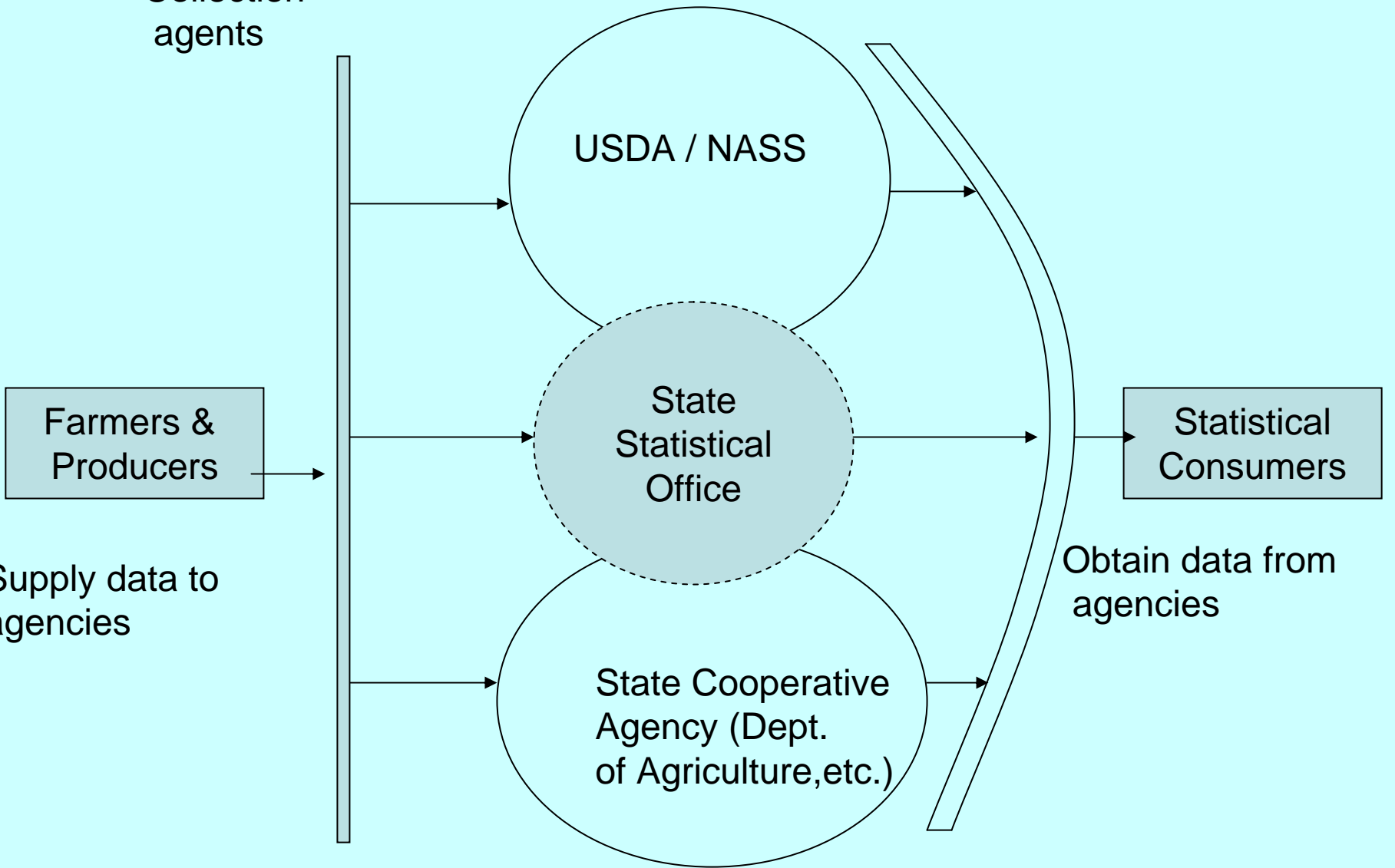


Semi-Supervised Approach



Vertical Integration: Agriculture

Collection agents



USDA / NASS

State
Statistical
Office

State Cooperative
Agency (Dept.
of Agriculture, etc.)

Farmers &
Producers

Statistical
Consumers

Obtain data from
agencies

Supply data to
agencies

Multiple Research Threads for the SKN

- Interfaces
- Metadata and Ontology
- Multi-leveled help
- Automatic slicing and dicing
- User needs and user testing
- Cross agency cooperation

- See www.ils.unc.edu/govstat

Library Partnerships

- Install and evaluate tools (e.g., RB) to add value to services
 - RB as alternative
 - RB as interface server
- Identify needs and resources
 - Specialized help
 - Local portals
- Triad partnerships: government agencies, special libraries, university researchers

Thank You!

Questions and Discussion
march@ils.unc.edu

www.ils.unc.edu/govstat