

Metadata for the SKN:
Philosophy, Progress, and Future
Directions

Sheila Denn, Dan Gillman, Carol
Hert, Jung Sun Oh, and Cristina
Pattueli

Metadata Philosophy

- To provide **sub-document level access** and **integration** across documents and agencies.
- To provide a **minimal set** of metadata elements necessary while allowing for **extensibility**.
- To achieve these goals in a manner that enables **efficient transfer to agencies**.

Progress to Date

- Prior to last status meeting:
 - Conducted a metadata user study to determine necessary elements from user perspective.
 - Started metadata modelling using Data Documentation Initiative (DDI) and ISO/IEC 11179 standards
- Since last status meeting
 - Developed a strategy to “test” and further develop the schema
 - Tested mark-up via a scenario
 - Through the markup process, determined that there was too much complexity in the data model for representing tabular data; developed a streamlined data model in response.

The Current Metadata Model

- Effort to balance complexity with functionality
 - Removal of elements designed to align data values and row/column headings with survey variables
- Retains ability to add on to the model to represent additional information using a “hierarchy of integration”

A Hierarchy of Integration

High level of
integration

- Linking of analysis units, universe statements, concept definitions, across documents and agencies
- Linking of row and column headings to underlying survey variables



*Our schema can
provide the items
beneath this dotted
line.*

Low level of
integration

- Linking of contextual information (such as footnotes) to tables, row/column headings, or data values
- Linking of data values to row and column headings
- Searchable row and column headings
- Searchable table titles

Our Schema in Action: An Example

- Scenario: The fact that the percentage of older people in the population of the US is increasing raises a question about the overall economic status of this group. In particular, we are interested in people who are **retired or no longer in the work force** and over a certain age (**65 or older**). We want to know the following things to understand the economic status of this particular group of people:
 - **Income level (in terms of median income) compared to the general (whole) population**
 - **Sources of income**
 - **Employment status**

Tables Identified to Respond to the Scenario

- Bureau of the Census

- Income Statistics

- (<http://www.census.gov/hhes/www/income.html>)

- Income in the United States 2002

- (<http://www.census.gov/prod/2003pubs/p60-221.pdf>)

- Table 3. Comparisons of Summary Measures of Money Income and Earnings by Selected Characteristics: 2001 and 2002

- » Markup available at

- <http://ils.unc.edu/govstat/metadata/table3census.xml>

- Table HINC-02. Age of Householder – Households by Total Money Income in 2002, Type of Household, Race, and Hispanic Origin of Householder

- (http://ferret.bls.census.gov/macro/032003/hhinc/new02_00.htm)

- Total, All Races

- (http://ferret.bls.census.gov/macro/032003/hhinc/new02_001.htm)

- Markup available at <http://ils.unc.edu/govstat/metadata/hinc02.xml>

Tables Identified to Respond to the Scenario (cont.)

- Social Security Administration
 - Social Welfare and the Economy, Annual Statistical Supplement, 2003, Poverty (3.E)
 - Table 3.E6. Percentage Distribution of Aged Families Receiving Social Security Benefits, by Share of Income from Benefits and Race, 2001
(<http://www.ssa.gov/policy/docs/statcomps/supplement/2003/3e.html>)
 - Income of the Population 55 or Older, 2000
 - Table 1.1. Percentage with Income from Specified Source, by Age, Marital Status, and Sex of Nonmarried Persons
(http://www.ssa.gov/policy/docs/statcomps/income_pop55/2000/sect1.html)
 - Markup available at
http://ils.unc.edu/govstat/metadata/SSA_Income_Source.xml

Tables Identified to Respond to the Scenario (cont.)

- Bureau of Labor Statistics
 - 3. Employment Status of the Civilian Noninstitutional Population by Age, Sex, and Race
(<ftp://ftp.bls.gov/pub/special.requests/lf/aat3.txt>)
 - 5. Employment Status of the Civilian Noninstitutional Population by Age, Sex, and Race
<ftp://ftp.bls.gov/pub/special.requests/lf/aat5.txt>
 - Markup available at
<http://ils.unc.edu/govstat/metadata/example5table5.xml>
 - Persons not in the Labor Force by Desire and Availability for Work, Age, and Sex
(<ftp://ftp.bls.gov/pub/special.requests/lf/aat35.txt>)

Examples from the Markup

- Table markup:
 - For each table, the schema encodes the **table title**, each **row** or **column heading**, and the **data values** in the table.
 - Each data value element references the row and column heading elements associated with it.
 - Footnotes are encoded at the highest level to which they apply – the table level, the row/column level, or the individual data value level.

Examples from the Markup (cont.)

Footnote that applies to the table as a whole is associated with the table title and can be displayed when the table as a whole is retrieved.

Footnote that applies only to a particular column or row is associated with the column or row and can be displayed when the column or row is retrieved.

Footnote that applies only to a particular data value is associated with the data value and can be displayed when the data value is retrieved.

```
<tableInfo>
  <tableTitle>Table 3. Comparison of Summary Measures of Money
Income and Earnings by Selected Characteristics: 2001 and
2002</tableTitle>
  <tableFootnote>Source: US Census Bureau, Current Population
Survey, 2002 and 2003 Annual Social and Economic
Supplements</tableFootnote>
  <tableFootnote>Households and people as of March of the
following year</tableFootnote>
  <rowInfo>
    <rowTitle>All households</rowTitle>
    <rowID>r001</rowID>
  ...
  <colInfo>
    <colTitle>2001 - Median money income - 90-percent
confidence interval</colTitle>
    <colFootnote>For an explanation of confidence
intervals, see "Standard Errors and Their Use" at
http://www.census.gov/hhes/income/income02/sa.pdf</colFootnote>
    <colFootnote>+/- dollars</colFootnote>
    <colID>c003</colID>
  </colInfo>
  ...
  <cellInfo>
    <cellValue rowID="r001" colID="c007">-
1.1</cellValue>
    <cellFootnote>Significantly different from zero at
the 90-percent confidence level</cellFootnote>
  </cellInfo>
```

Examples from the Markup (cont.)

Each row and column has a unique identifier.

```
<tableInfo>
  <tableTitle>Table 3. Comparison of Summary Measures of Money
Income and Earnings by Selected Characteristics: 2001 and
2002</tableTitle>
  <tableFootnote>Source: US Census Bureau, Current Population
Survey, 2002 and 2003 Annual Social and Economic
Supplements</tableFootnote>
  <tableFootnote>Households and people as of March of the
following year</tableFootnote>
  <rowInfo>
    <rowTitle>All households</rowTitle>
    <rowID>r001</rowID>
    ...
  <colInfo>
    <colTitle>2001 - Median money income - 90-percent
confidence interval</colTitle>
    <colFootnote>For an explanation of confidence
intervals, see "Standard Errors and Their Use" at
http://www.census.gov/hhes/income/income02/sa.pdf</colFootnote>
    <colFootnote>+/- dollars</colFootnote>
    <colID>c003</colID>
  </colInfo>
  ...
  <cellInfo>
    <cellValue rowID="r001" colID="c007">-
1.1</cellValue>
    <cellFootnote>Significantly different from zero at
the 90-percent confidence level</cellFootnote>
  </cellInfo>
```

Each data value contains a reference to the particular row/column combination with which it is associated.

Examples from the Markup (cont.)

Percentage with income from specified source, by age, marital status,

Source of income

Earnings

Wages and salaries

Self-employment

Retirement benefits

Social Security ^a

Benefits other than Social Security

Other public pensions

Railroad Retirement

Government employee pensions

Military

Federal

State or local

Private pensions or annuities

In order to preserve category information, individual row and column headings include the category labelling.

Including the category labelling within the row/column headings improves access to data embedded within tables by making the category information searchable.

```
<tableInfo>
  <tableTitle>Table 1.1 Percentage with income from
  specified source, by age, marital status, and sex
  of nonmarried persons</tableTitle>
  <rowInfo>
    <rowTitle>Source of Income -
    Earnings</rowTitle>
    <rowID>r001</rowID>
  </rowInfo>
  <rowInfo>
    <rowTitle>Source of Income -
    Earnings - Wages and salaries</rowTitle>
    <rowID>r002</rowID>
  </rowInfo>
  <rowInfo>
    <rowTitle>Source of Income -
    Earnings - Self-employment</rowTitle>
    <rowID>r003</rowID>
  </rowInfo>
  <rowInfo>
    <rowTitle>Source of Income -
    Retirement benefits</rowTitle>
    <rowID>r004</rowID>
  </rowInfo>
  <rowInfo>
    <rowTitle>Source of Income -
    Retirement benefits - Social Security</rowTitle>
    <rowFootnote>Social Security
    includes retired-worker benefits, dependents' or
    survivors' benefits, disability benefits,
    transitionally insured benefits, or
    special age-72 benefits</rowFootnote>
    <rowID>r005</rowID>
  </rowInfo>
  ...

```

Examples from the Markup (cont.)

Table 1.1
Percentage with income from specified source, by age, marital status, and sex of nonmarried persons

Source of income	Aged 55-61	Aged 62-64	Aged 65 or older					
			Total	65-69	70-74	75-79	80-84	85 or older
Earnings	80	64	22	44	26	14	7	4
Wages and salaries	76	59	19	39	22	12	6	3

```

<tableTitle>Table 1.1 Percentage with income from specified source, by age, marital status,
  and sex of nonmarried persons</tableTitle>
<colInfo>
  <colTitle>Aged 65 or older Total All units</colTitle>
  <colID>c003</colID>
</colInfo>

<rowInfo>
  <rowTitle>Source of Income - Earnings - Wages and salaries</rowTitle>
  <rowID>r002</rowID>
</rowInfo>

<cellInfo>
  <cellValue rowID="r002" colID="c003">19</cellValue>
</cellInfo>
  
```

Examples from the Markup (cont.)

TABLE 3.
Comparison of Summary Measures of Money Income and Earnings by Selected Characteristics: 2001 and 2002

(Households and people as of March of the following year)

Characteristic	2001			2002		
	Number (thousands)	Median money income (in 2002 dollars)		Number (thousands)	Median money income	
		Value (dollars)	90-percent confidence interval ¹ (± dollars)		Value (dollars)	90-percent confidence interval ¹ (± dollars)
HOUSEHOLDS						
All households	109,297	42,900	216	111,278	42,409	229
Type of Household						
Family households	74,329	53,106	294	75,596	52,704	396
Married-couple families	56,747	61,433	348	57,320	61,254	327
Female householder, no husband present	13,143	28,590	483	13,620	29,001	497
Male householder, no wife present	4,438	41,363	874	4,656	41,711	684
Nonfamily households	34,969	26,039	282	35,682	25,406	281
Female householder	19,390	20,586	353	19,662	20,913	316
Male householder	15,579	32,826	401	16,020	31,404	433
Age of Householder						
Under 65 years	86,821	50,010	333	88,619	49,510	352
15 to 24 years	6,391	28,644	812	6,611	27,828	748
25 to 34 years	18,988	45,797	623	19,055	45,330	484
35 to 44 years	24,031	54,168	700	24,069	53,521	679
45 to 54 years	22,208	58,968	814	22,623	59,021	864
55 to 64 years	15,203	46,593	710	16,260	47,203	702
65 years and over	22,476	23,486	319	22,659	23,152	309

```

<tableTitle>Table 3. Comparison of
Summary Measures of Money Income
and Earnings by Selected
Characteristics: 2001 and
2002</tableTitle>

<tableFootnote>Source: US Census Bureau,
Current Population Survey, 2002 and
2003 Annual Social and Economic
Supplements</tableFootnote>

<tableFootnote>Households and people as
of March of the following
year</tableFootnote>

<rowInfo>
<rowTitle>Age of Householder - 65
years and over</rowTitle>
<rowID>r015</rowID>
</rowInfo>

<colInfo>
<colTitle>2002 - Median money
income - value</colTitle>
<colFootnote>dollars</colFootnote>
<colID>c005</colID>
</colInfo>

<cellInfo>
<cellValue rowID="r015"
colID="c005">23,152</cellValue>
</cellInfo>

```

Examples from the Markup (cont.)

Table 1.1
Percentage with income from specified source, by age, marital status, and sex of nonmarried persons

Source of income	Aged 55-61	Aged 62-64	Aged 65 or older					
			Total	65-69	70-74	75-79	80-84	85 or older
Earnings	80	64	22	44	26	14	7	4
Wages and salaries	76	59	19	39	22	12	6	3

```

<colInfo>
  <colTitle>Aged 65 or older Total All units</colTitle>
  <colID>c003</colID>
</colInfo>
<rowInfo>
  <rowTitle>Source of Income - Earnings - Wages and salaries</rowTitle>
  <rowID>r002</rowID>
</rowInfo>
<cellInfo>
  <cellValue rowID="r002" colID="c003">19</cellValue>
</cellInfo>
  
```

Table 3.
Comparison of Summary Measures of Money Income and Earnings by Select Characteristics: 2001 and 2002

(Households and people as of March of the following year)

Characteristic	2001			2002		
	Households	People	Percentage	Households	People	Percentage
Age of Householder						
Under 65 years						
15 to 24 years						
25 to 34 years						
35 to 44 years						
45 to 54 years						
55 to 64 years						
65 years and over	22,476	23,486	319	22,659	23,152	309

Note that since these headings both contain keywords for age 65 or older that we can begin to think about ways to integrate these data.

```

<rowInfo>
  <rowTitle>Age of Householder - 65 years and over</rowTitle>
  <rowID>r015</rowID>
</rowInfo>
<colInfo>
  <colTitle>2002 - Median money income - value</colTitle>
  <colFootnote>dollars</colFootnote>
  <colID>c005</colID>
</colInfo>
<cellInfo>
  <cellValue rowID="r015" colID="c005">23,152</cellValue>
</cellInfo>
  
```

What the Example Demonstrates

- Access: preserving data from table titles, row/column headings, and footnotes allows metadata essential for understanding to travel with the data values, and aids in search and retrieval
- Integration: once we have this essential metadata tagged, it becomes easier to use tag similarities to allow us to investigate options for displaying data from different tables in an integrated manner.

We Need Your Help!

Discussion Points for May 14, 2004

Topic 1: Do we have the right elements for your needs? Can you get the necessary info to fill the elements?

Topic 2: What metadata initiatives are in action in your organization that we need to map to?

Topic 3: What are the ways in which we can partner to collect the necessary metadata? What is a reasonable level of effort on the agency side to support this metadata model? What obstacles are there? How can we go about working with you to develop a training program to implement this model?

Related Materials

- Current schema model:
http://ils.unc.edu/govstat/metadata/govstat_schema.xml
- Developing an SKN Metadata Model: Statement of Work:
http://ils.unc.edu/govstat/papers/proposal_metadata_modelling.doc
- Integration Example (Economic status of aged people):
http://ils.unc.edu/govstat/papers/Scenario_UNC_1.doc
- Metadata to Support comparisons example:
http://ils.unc.edu/govstat/papers/comparison_scenarios.doc