

## Building a Metadata Repository to Support the 2002 Economic Census



Samuel N. Highsmith  
U.S. Census Bureau  
Samuel.N.Highsmith.Jr@census.gov  
301-457-1928

USCENSUSBUREAU  
Helping You Make Informed Decisions

## Agenda



- ↖ Background on the Corporate Metadata Repository project
- ↖ Building an Economic Pilot Application
- ↖ Building the Economic Metadata Repository
- ↖ Lessons Learned
- ↖ Future Plans

USCENSUSBUREAU  
Helping You Make Informed Decisions

## Why a CMR?



### ↖ If you have:

- Documentation
- Descriptions of processes
- Program definitions
- Code book
- Data flow diagrams

### ↖ You already have metadata

### ↖ The CMR helps you organize, share, reuse

USCENSUSBUREAU

Helping You Make Informed Decisions

## CMR Background



### ↖ Presented in ISIS 2000 paper at Riga, Latvia

### ↖ Based on survey and census life cycle model

- Design
- Collection
- Editing and Processing
- Data Dissemination

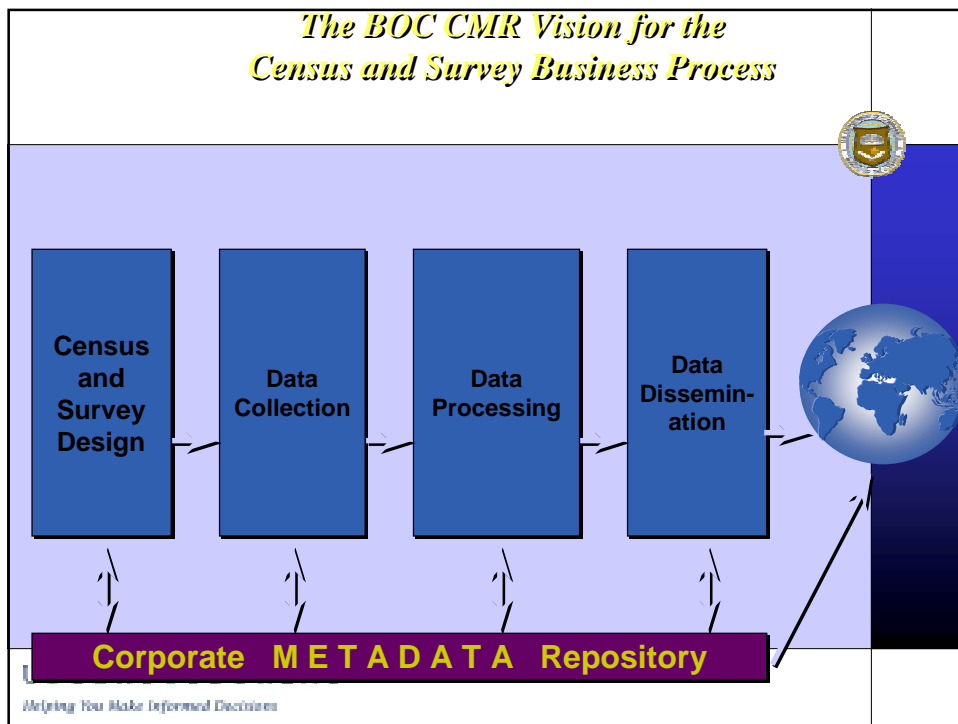
### ↖ Card Catalog for Survey Processing

### ↖ Metadata Provider for Many Systems - XML

USCENSUSBUREAU

Helping You Make Informed Decisions

## The BOC CMR Vision for the Census and Survey Business Process



## Corporate Metadata Repository Background

### Business Data Model

- Developed in conjunction with work from Sweden, Canada, Australia, UNECE workshops
- Reinvention Lab With Many Participants
- Survey and Census business model

### Data Element Registry

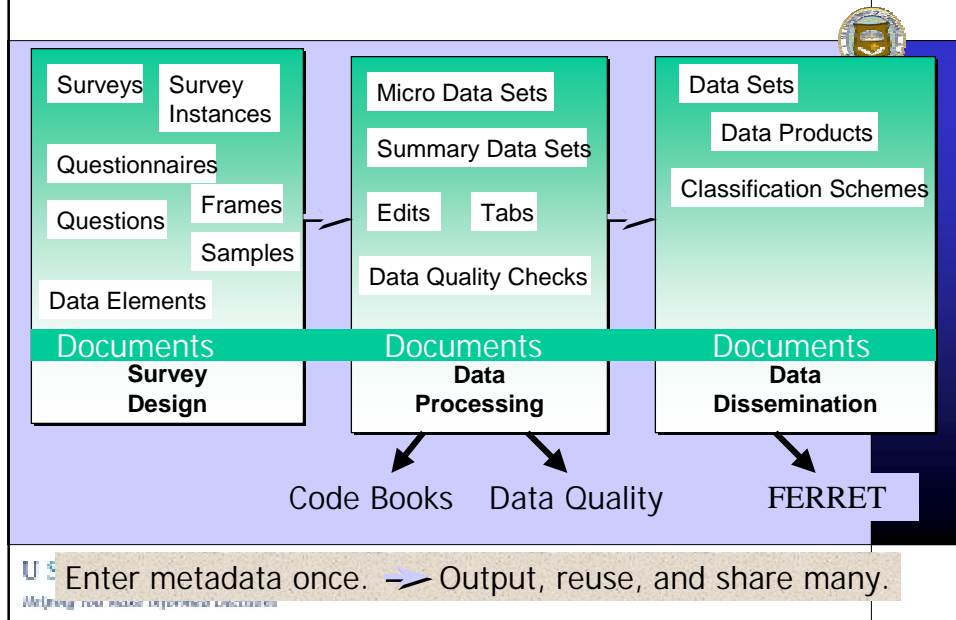
- Model built for us under contract
- Implements ISO 11179

### Combined BDM and DER model to produce the CMR

USCENSUSBUREAU

Helping You Make Informed Decisions

## Maintain & Use Structured & Unstructured Metadata



## Corporate Metadata Repository Current Users

- American Factfinder
  - All Geography Maps after June
- Economic Directorate for 2002 Economic Census - CMR and DER
- Web Portals
  - CMR
  - Project Management
  - DSD for CPS/SIPP Documentation
- DER at other agencies - [www://fdr.faa.gov](http://www.fdr.faa.gov)

USCENSUSBUREAU

Helping You Make Informed Decisions

U.S. Census Bureau  
**CMR**  
 Composite Metadata Repository

August 18, 2001

Search:

Home | Surveys | Data Elements | Products | Data Quality | Bus Rules | Admin

Application | Reports | Interchanges | Documentation

**Survey Access**  
 Info  
 Survey  
 Survey Instances  
 Data Sets  
 Questionnaires  
 Products

**Survey Links**  
 American Community Survey (ACS)  
 American FactFinder  
 Source for Census 2000 and other  
 Census Bureau census/survey data.  
 Current Population Survey (CPS)

**Survey Area**  
 CDR Overview

**Available Components**  
 Surveys |  
 Survey Instances |  
 Datasets |

Other Census Portals  
 Current Population Survey (CPS)  
 Project Management Repository (PMR)  
 Quality Management Repository (QMR)

Help Desk  
 Feedback  
 Chat

USCENSUSBUREAU  
 Helping You Make Informed Decisions

U.S. Bureau  
**CMR**  
 Composite Metadata Repository

August 18, 2001

Search:

Home | Surveys | Data Elements | Products | Data Quality | Bus Rules | Admin

Application | Reports | Interchanges | Documentation

**CDR Access**  
 Info  
 Data Elements  
 Data Concepts  
 Value Domains  
 Conceptual Domains  
 Classification Schemes  
 Stewards  
 Submissions  
 Regulations

**CDR Links**  
 DDC/DC 111/3 Standard 111/2  
 Implementation Guidelines - FactData 111/3  
 Data Registry - LNA Data Registry - US  
 Health Info. Knowledgebase

**HTML Portal**  
 CDR Overview  
 CDR Folder

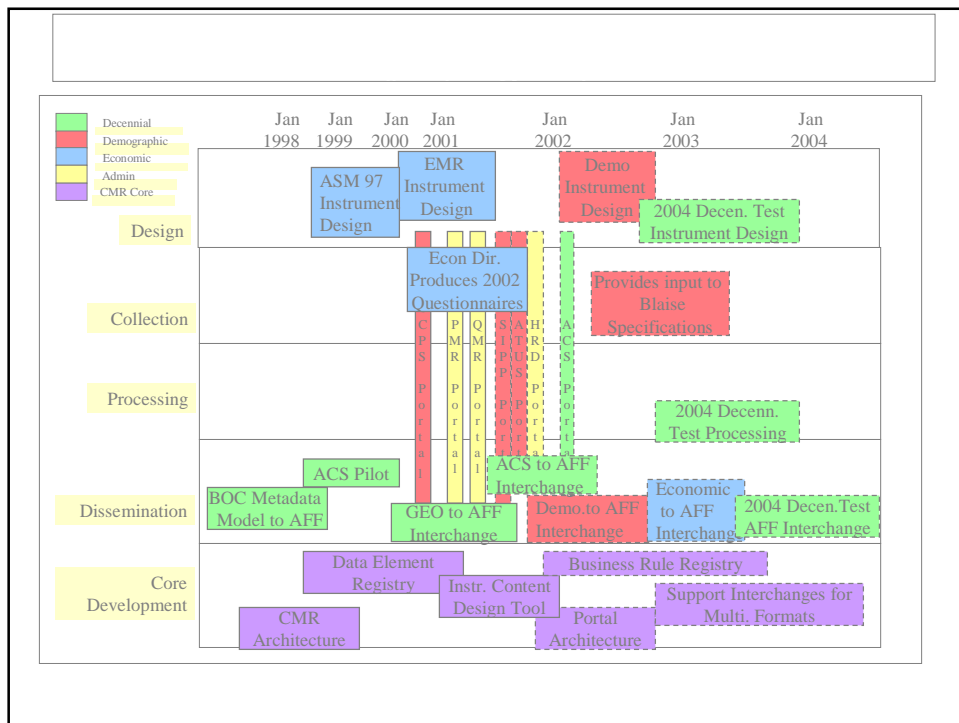
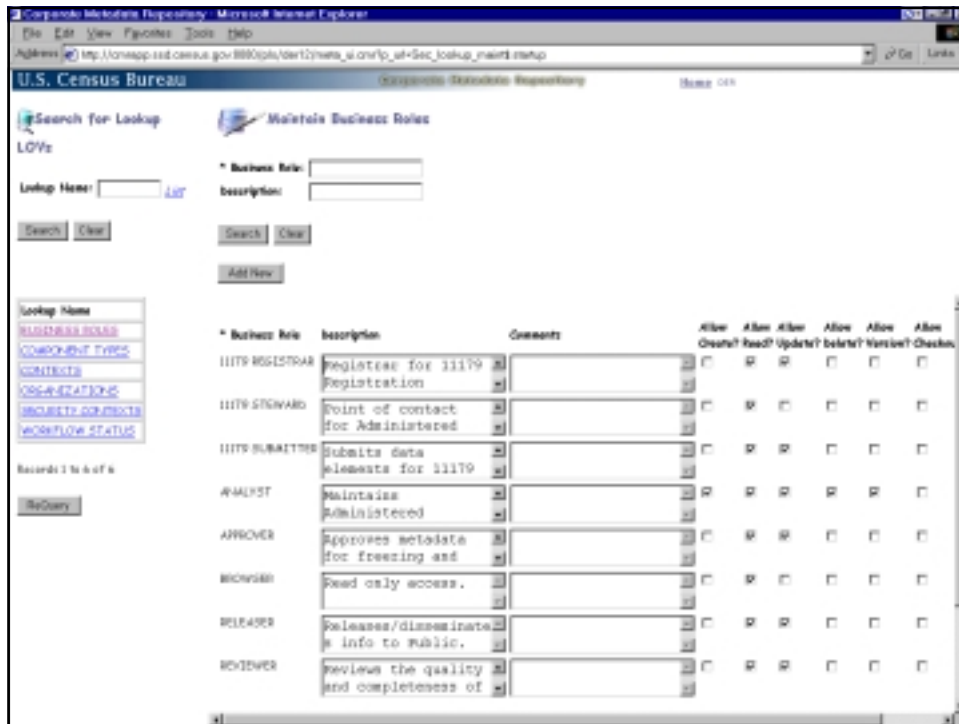
**Available Components**

Data Elements	1,201
Data Element Concepts	13
Value Domains	873
Conceptual Domains	1
Classifications	6

Other Census Portals  
 Current Population Survey (CPS)  
 Project Management Repository (PMR)  
 Quality Management Repository (QMR)

Help Desk  
 Feedback  
 Chat

USCENSUSBUREAU  
 Helping You Make Informed Decisions



## CMR Core Development

Module #1	% Complete
- CMR Architecture *	100%
- Data Element Registry *	100%
- Data Set Registry	50%
<b>Module #2 Survey Design</b>	
- Questionnaire Registry *	80%
- Question Registry *	70%
- Instrument Content Design *	90%
- Survey Registry	70%
- Survey Instances Registry	70%
- Frame Registry	0%
- Sample Registry	0%
- Universe Registry	0%
- Business. Rule Registry	20%
<b>Module #3 Processing &amp; Dissemination</b>	
- SAS Interchange	0%
- Tier 1 Product Registry for AFF	40%
- Associated Documentation	40%
- Validate Metadata for AFF	25%
- Support Multiple Formats	20%

\* - denotes components currently available for use

## Build a Pilot Application for the Economic Directorate

- Proof of Concept
- Provided Value Added to existing Metadata repositories
- C/R/U/D tool for their Data Dissemination Metadata system
- Tool for Comparability for MCD
- Many JAD Sessions and Six Months to Build In 1999
- Result - Store All Reusable Metadata in EMR

USCENSUSBUREAU

Helping You Make Informed Decisions

## Economic Census 2002



### EMR Provides

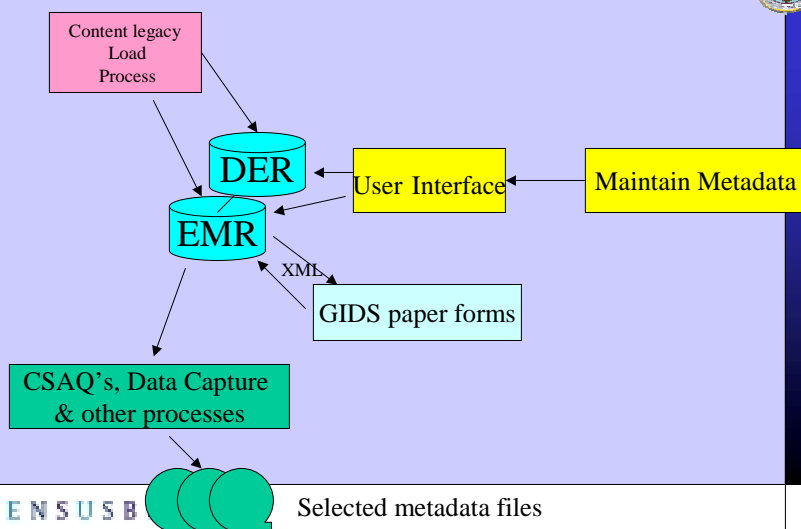
- Security
- Legacy Load
- Implements Business Rules
- Directly Tied to Data Element Registry

### Metadata Provider to the GIDS software

USCENSUSBUREAU

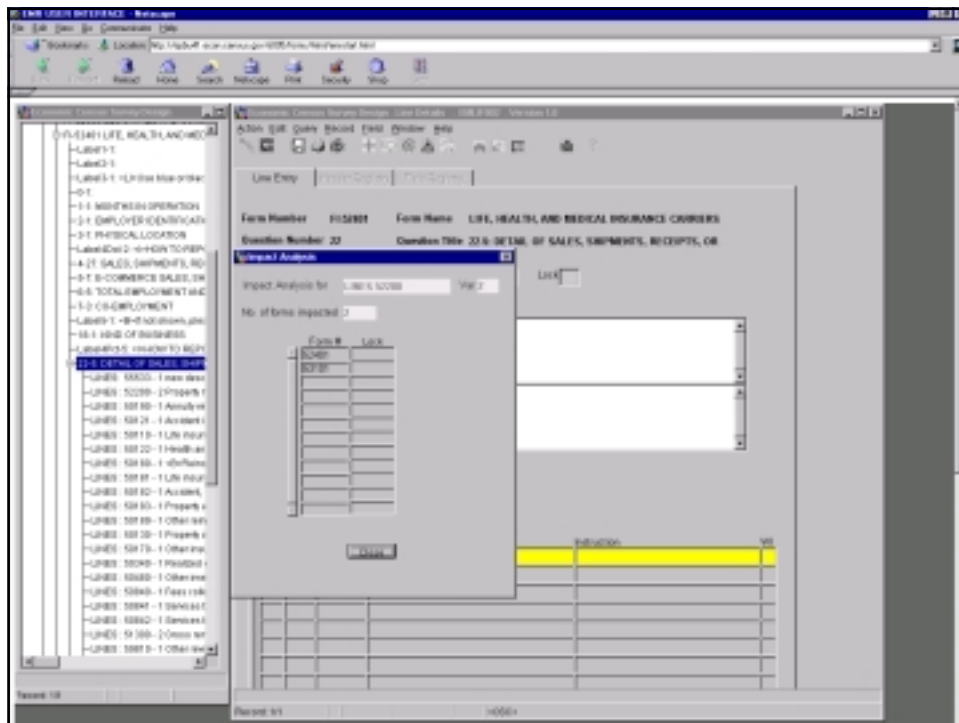
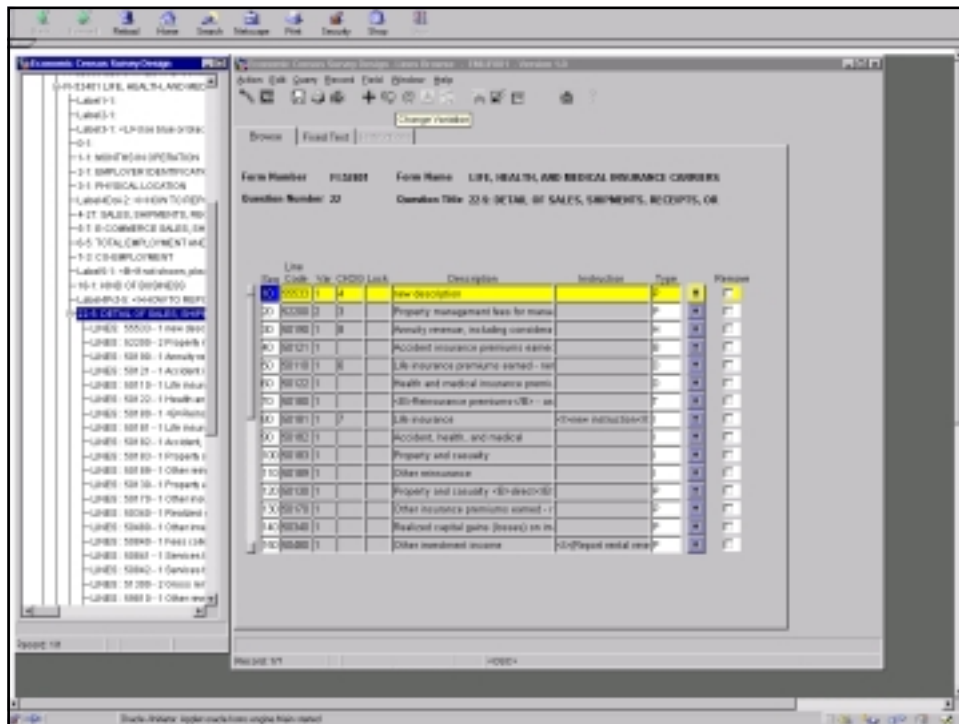
Helping You Make Informed Decisions

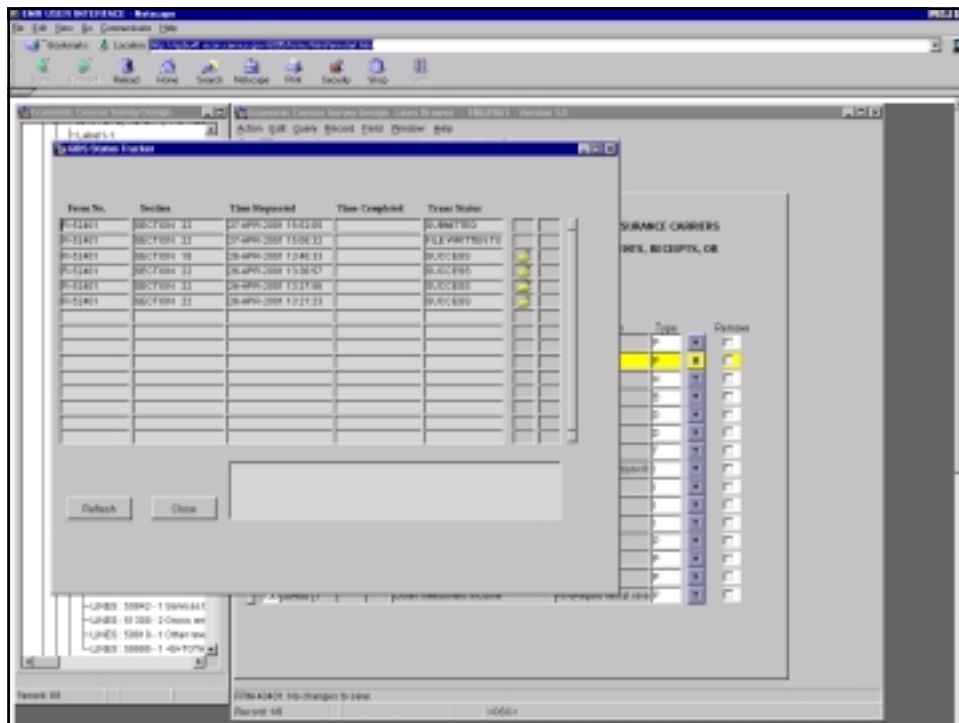
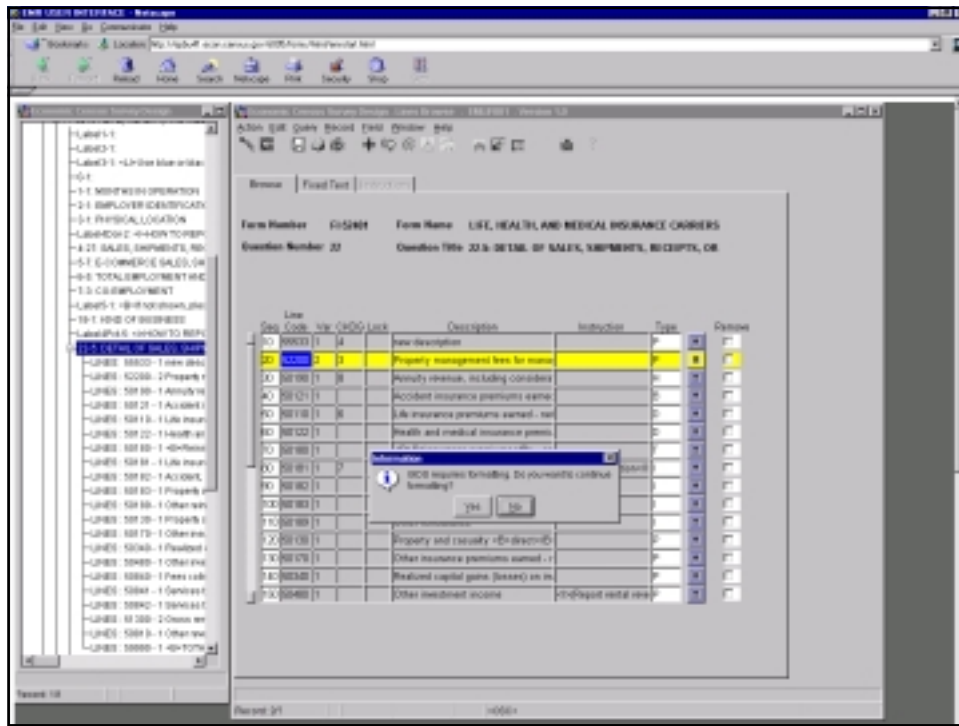
## Economic Metadata Repository Architecture

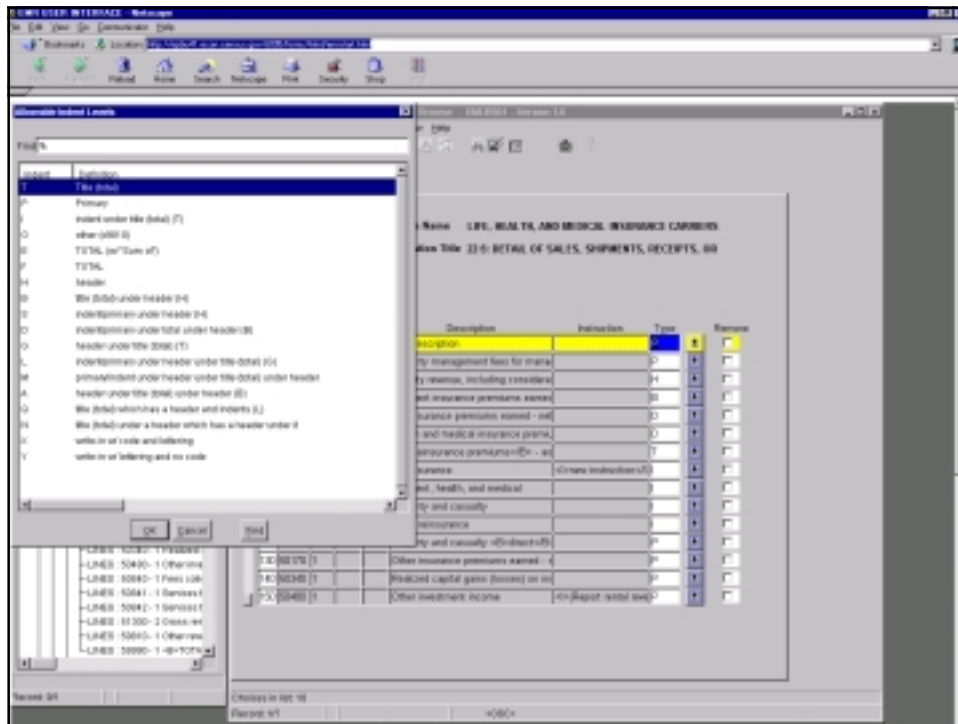










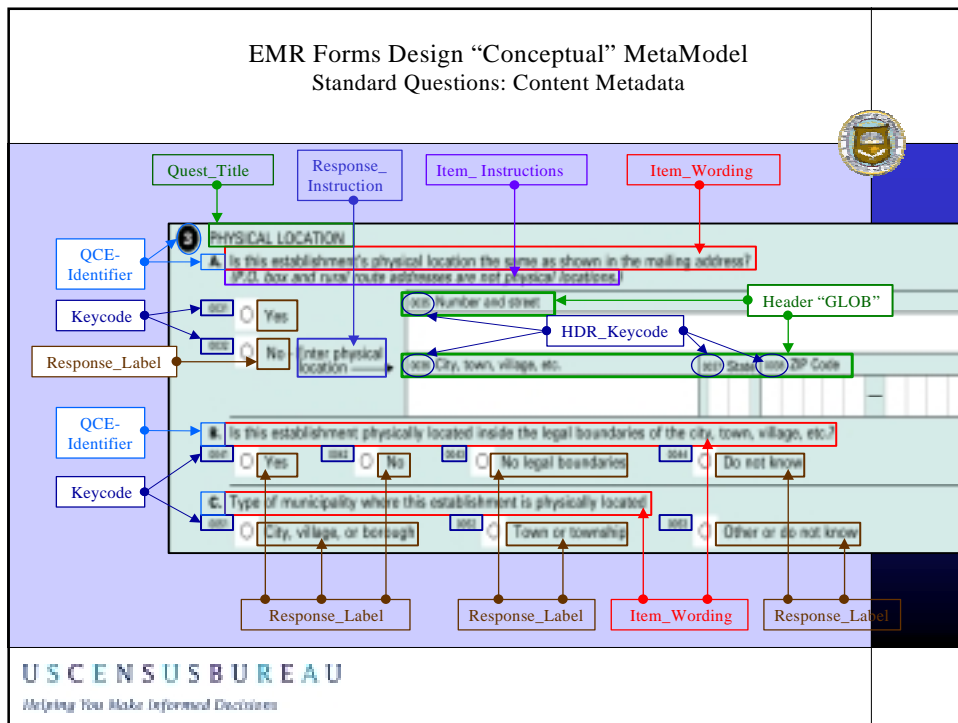
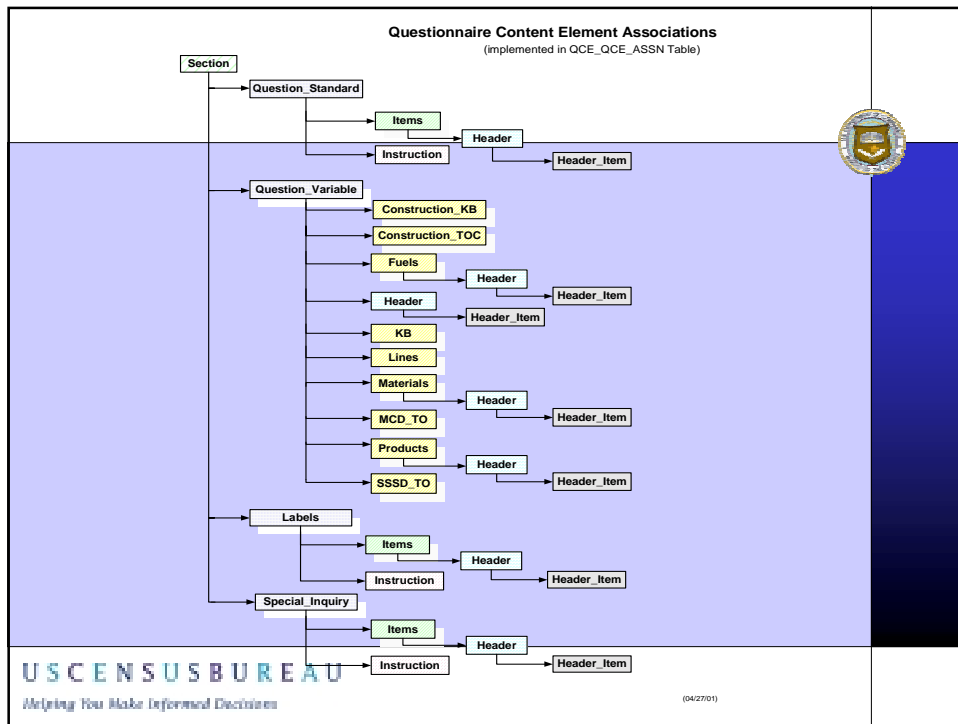


## XML Interchange

- Autoformatted Section
- Variable Content Section
- Standard Question Section
- Headers
- Instructions
- Item Groups
- Forms Design and Assembly
- Returned Image Files feed Viewer

USCENSUSBUREAU

Helping You Make Informed Decisions



## EMR to GIDS XML Interchange Standard Question 19: Item Rows & Data Elements



**10 CLASS OF CUSTOMER**  
(On-line information access providers only.)  
Estimates are acceptable.

**A.** Did this establishment sell (retail or wholesale) to the general public?

**B.** Did you sell ONLY to retailers or wholesalers for resale?

**C.** Did you require special qualifications (e.g., proof of business or professional license, tax ID, commercial account) from customers in order to make sales?

Item Row # 1 points to the question header.

Item Row # 2 points to the 'Yes' radio button for question A.

Item Row # 3 points to the 'Yes' radio button for question B.

Item Row # 4 points to the 'Yes' radio button for question C.

Data Elements are indicated by arrows pointing to the 'Yes' radio buttons for questions A, B, and C.

Cont...

**US CENSUS BUREAU**  
Helping You Make Informed Decisions

## EMR to GIDS XML Interchange Standard Question 19: Item Rows & Data Elements



**10 CLASS OF CUSTOMER - Continued**

**D.** Percentage of this establishment's total sales reported in **10** line A) for each class of customer

	2002 Whole percent of sales and receipts
<b>1.</b> Household consumers and individual users	000
<b>2.</b> Export sales	000
<b>3.</b> Retailers for resale	000
<b>4.</b> Other wholesale establishments for resale	000
<b>5.</b> Repair shops for use in repair work	000
<b>6.</b> Industrial users for use in production (manufacturing and mining)	000
<b>7.</b> Restaurants, hotels, food services, and contract feeding	000
<b>8.</b> Businesses for end use, not for resale or production	000
<b>9.</b> Building contractors, heavy construction, and special trade contractors	000
<b>10.</b> Farmers for use in farm production	000
<b>11.</b> Governmental bodies (Federal, State, and local)	000
<b>12.</b> Other - Specify	000
<b>13. TOTAL</b>	100

Item Row # 5 points to question D.

Item Row # 6 points to line 1.

Item Row # 7 points to line 2.

Item Row # 17 points to line 12.

Item Row # 18 points to the 'Specify' field for line 12.

Item Row # 19 points to line 13.

Data Elements are indicated by arrows pointing to the percentage columns for lines 1, 2, 12, and 13.

**US CENSUS BUREAU**  
Helping You Make Informed Decisions

## Forms Design Background



### ↖ Prior Censuses: 3 different styles of forms

- Census of Construction
- Censuses of Manufacturing and Mining
- Services censuses

### ↖ Goals For 2002

- Consistent, standardized wording, question placement, definitions, data elements
- Offer multiple response mechanisms

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

Helping You Make Informed Decisions

## 2002 Economic Census has about 650 Variations of the questionnaire



### ↖ Previously All Pages Individually Coded

### ↖ Sent to Publications Division for Preparation

### ↖ 1-3 Months per form

### ↖ Rampant Duplication

### ↖ Many Equivalent Variations of Questions

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

Helping You Make Informed Decisions

## Services Division Forms



### ↳ About 199 Variations of the Questionnaire

- Approximately 1700 Pages
- About 132 Custom standard section variations (manually designed)
- About 100 Custom special inquiry variations (manually designed)
- 3 Autoformatted sections (50-75%) (Using predefined rules)

USCENSUSBUREAU

Helping You Make Informed Decisions

## Key benefits of differentiating between content and layout



- ↳ Multiple layouts can be associated with one content (example: electronic and paper)
- ↳ Changes to content automatically flow into layout via software process
- ↳ Layout characteristics can be changed without changing content
- ↳ Many layouts can be automatically generated

USCENSUSBUREAU

Helping You Make Informed Decisions



### Content (EMR)

Sig	NACS	Var	CKDG	Lock	Description	Instruction	KDHead	Remove
1D	5200020	1	1		Direct insurance carriers		1	<input type="checkbox"/>
2D	52411310	1	9		Life insurance carrier			<input type="checkbox"/>
3D	52411310	2	7		Fraternal life insurance organization			<input type="checkbox"/>
4D	52411390	1	2		Accident and disability income insurer			<input type="checkbox"/>
5D	52411410	1	0		Health insurance carrier			<input type="checkbox"/>



### Content with Layout (GIDS)

**18. KIND OF BUSINESS**  
Principal kind of business in 2002  
(Mark "X" only ONE box.)

**Direct insurance carriers**

524 113 19 19  Life insurance carrier

524 113 19 27  Fraternal life insurance organization

524 113 99 12  Accident and disability income insurance carrier

524 114 19 18  Health insurance carrier

USCENSUSBUREAU

Helping You Make Informed Decisions

## Same content, different layouts, same data elements



Products and services				Census product code	Unit of measure for quantities	Products shipped and other receipts			
(A)				(B)	(C)	Quantity	Value, f.o.b. plant		
						(D)	(E)		
							Millions	Thous-	Dollars
							sands		
<b>Ice Cream and Ices</b>	ICE CREAM (Including custards)	Ice cream (excluding lowfat and nonfat)	Shipped in bulk (containers 3 gallons or more)	311 520 0111 5	Thousand gallons				
			Shipped in container sizes (less than 3 gallons)	311 520 0221 6					
			Novelty forms	311 520 0331 7					

Products and services				Census product code	Unit of measure for quantities	Products shipped and other receipts			
(A)				(B)	(C)	Quantity	Value, f.o.b. plant		
						(D)	(E)		
							Millions	Thous-	Dollars
							sands		
<b>43. Ice cream and ices</b>	Ice cream, including custards	Ice cream, excluding lowfat and nonfat	Shipped in bulk (containers 3 gallons or more)	311 520 0111 5	Thousand gallons				
			Shipped in container sizes (less than 3 gallons)	311 520 0221 6					
			Novelty forms	311 520 0331 7					

USCEN

Helping You Make Informed Decisions

## Content formatted into a forms layout

Content

Line No.	Products and services	Comms product code	Unit of measure by quantities	Quantity produced for all members	Products shipped and other receipts		
					Quantity	Value, \$ Bill. plant	Thous.
1	Sugar cane (with) products and byproducts Raw cane sugar .....	311 311 0111 1					
2	Other sugar cane mill products and byproducts Molasses and syrup, including sugar cane and Molasses, except refinery' molasses	311 311 0011 2					
3	All other sugar cane mill products and byproducts, including sugar for consumer use without further processing, bagasse, etc.	311 311 0011 3					

Shading

Lines

Graphic Elements

Answer Cells  
(data elements)

USCENSUSBUREAU

Helping You Make Informed Decisions

## Lessons Learned

- Metadata quality critical
- Build a Metadata import tool
- Migration of legacy metadata is a large undertaking
- Re-engineering is learning on the job
- Can't sell infra-structure
- Fully web enabled applications force organizational change
- Teamwork = Users working with Developers

USCENSUSBUREAU

Helping You Make Informed Decisions

## Future Plans



- ↪ Data Capture - Biggest cost savings
- ↪ Electronic Instruments
- ↪ Dissemination support EMR to AFF
- ↪ Expanded data cleaning for CMR
- ↪ XML schema using XSL for translation

USCENSUSBUREAU  
Helping You Make Informed Decisions

## End Results



- ↪ Produced 200 final forms for Services Division
- ↪ Manufacturing and Construction Division Forms - March through September
- ↪ CMR - Infrastructure and Applications

USCENSUSBUREAU  
Helping You Make Informed Decisions



**Questions?**

**USCENSUSBUREAU**

Helping You Make Informed Decisions