Using and Developing Survey Metadata

Debra Reed-Gillette - dsr3@cdc.gov
Michael Volynski, Ph.D. - mhv3@cdc.gov

2006
Agenda

- Who we are
- Survey creation: process and roles
- Metadata key features
- Metadata hierarchy
- Metadata attributes
- Metadata XML structure
- Meta Editor – authoring tool
What is NHANES?

• The National Health and Nutrition Examination Survey (NHANES) is one of several Centers for Disease Control and Prevention (CDC) surveys used for assessing the health of the United States population.
• The primary objective of NHANES is to collect high quality health and nutrition data and to release it for public use in a timely manner.
NHANES

- 5000 Participants
- 15 Cities / year
- All ages
- 12 months / year
- HH screening and interview
- Mobile Exam Center exam, interview, blood draw
NHANES
Typical Survey Creation Roles

• Content expert
  • Content design: specifications, protocols, other documents

• Content developer
  • Content creation: codebooks, metadata
  • Content maintenance

• Software developer
  • Survey software development
  • Survey software deployment
  • Survey software maintenance
CHITA Main Concepts

• Clear separation between survey content (creation and management) and survey software (creation and management)
• Content expert plays dominant role
• Software developer plays minimized role
• Using WYSIWYG visual editor to create content
• Content exists as Metadata in metadata repository
CHITA Survey Creation Roles

• Content expert
  • Content design: specifications, protocols, other documents
    • Content maintenance
• Content developer
  • Content creation: codebooks, metadata
  • Content maintenance
    • Content deployment
    • Survey maintenance
• Software developer
  • Survey tools development
  • Survey software development
  • Survey software deployment
  • Survey software maintenance
Metadata Usage

- Data Collection
- Document Production
- Data Delivery
- Metadata Exchange
- Data Analysis
Metadata Use

- Metadata describes survey content
- Metadata created via visual tool
- Metadata resides in metadata repository in relational database and in XML
- Metadata used for:
  - Data collection
  - Data delivery
  - Data analysis
  - Codebooks, specifications, protocols, and other documents
  - Metadata exchange
Metadata Driven Components

- Questionnaires:
  - CAPI
  - ACASI
  - PDA
  - WEB

- Examination/biologic components:
  - Measurement components (BM, BP, …)
  - Specimen components (UC, PH, …)
  - Environmental components (Air, Water, …)
Metadata Main Features

- Hierarchy: survey/module/section/item
- Text: questions, instructions, codes, helps
- Media: sounds, images, video
- Flow: skips, branches, check items
- Validations and custom errors
- Presentation for data collection
- Gender/Proxy/Data dynamic substitutions
- Storage for data collection
- Data analysis / SAS generation
Data Collection Features

• Windows, PDA or WEB front-end
• Rich GUI
• Sound, image and video support
• Phonetic multi-column dictionary support
• Local and/or remote data storage with synchronization
Meta Editor

• Visual Tool to create Metadata
• Oriented toward content experts and power users
• Insert/Delete/Copy/Paste
• Templates for common tasks
• Validation on the fly
• Item data and item flow preview
• Maintains master repository with export/import
• Maintains surveys
• Generates XML to drive data collection and data analysis
Metadata Key Features

- Sophisticated hierarchical data structure
- Built-in version control
- Stored in relational database
- No limits on number of modules, sections, subsections, nodes
- Multiple node and item types
- 50+ node and item attributes
- Sub items support
- Media file support
- XML import/export support
Metadata Key Features II

- Multiple target support
- Multiple code/response support
- Sophisticated data collection validation
- Data collection presentation
- Skip pattern and check item support
- Substitution pattern support
- Data delivery formats
- Shared between CAPI, ACASI, CIDI, …
Metadata Hierarchy

• Repository root
  • Survey 1
    • Survey 2
      • Component 1
        • Module 1
        • Module 2
        • ...
      • Component 2
        • ...
  • Survey 3
    • ...
  • ...
  • ...
Metadata Hierarchy II

- Module1
  - Section1
    - Item1
      - Subitem 1
      - Subitem 2
    - Item2
    - ...
  - Section2
    - Item1
    - Item2
    - ...
    - ...
- Module2
  - ...
  - ...

(CDC logo)
## Meta Editor: Master Repository

![Metadata Editor interface](image)

<table>
<thead>
<tr>
<th>Name</th>
<th>Node Type</th>
<th>Description</th>
<th>Targets</th>
<th>Item ID</th>
<th>Item ...</th>
<th>SAS Label</th>
<th>Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>MI-BLAISE</td>
<td>Module</td>
<td>Blaise Module</td>
<td>Both, 8 Years - 150 Years</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MI-CAPI</td>
<td>Module</td>
<td>CAPI Module</td>
<td>Both, 12 Years - 150 Years</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MI-ACASI</td>
<td>Module</td>
<td>Audio CASI ...</td>
<td>Both, 12 Years - 59 Years</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MI-CDI</td>
<td>Module</td>
<td>Critical Data ...</td>
<td>Both, 12 Years - 150 Years</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**General Properties (ID: 11026; Index: 0)**

- **Sequence #**: 5
- **Name**: MI-05
- **Type**: Survey
- **Description**: NHANES MI-05

**Database Properties**

- **Collection Item ID**
- **Analytic Item ID**
- **Database Table**
- **Database Type**

**Item Properties**

- **Item Type**
- **Lookup Field**
- **SAS Type**
- **Length**
- **Label**
<table>
<thead>
<tr>
<th>General</th>
<th>Targets</th>
<th>Text</th>
<th>Codes</th>
<th>Help</th>
<th>Versions</th>
<th>Preview</th>
</tr>
</thead>
</table>

**English Text:**
During your life, on how many days have you had at least one drink of alcohol?

**English Instructions:**
Please select one of the following choices.

**Spanish Text:**
Durante su vida, ¿cuántos días ha tomado al menos un trago de alcohol?

**Spanish Instructions:**
Por favor seleccione una de las siguientes opciones.

**CAPI Instructions**

**Sounds**
- English: C:\CHETA05\NCSACASI\NHANES_Sounds\Sounds\English\ALQ\alq022.wav
- Spanish: C:\CHETA05\NCSACASI\NHANES_Sounds\Sounds\Spanish\ALQ\alq022.wav
What is your maiden name?

VERIFY SPELLING.

- Enter maiden name
- Same as current last name Smith
Metadata XMLs

- Component/questionnaire content XML
- Substitution patterns XML
- Cross-component media files XML
- Multi-column dictionary (s) XML(s)
Metadata XML Structure
Content XML Items

- <Subnodes>
  - <Node ID="10982" Name="SMQ620_" Path="NHANES-05\MI-05\MI-ACASI\SMQ\SMQ620_" Type="4" Description="" Index="0" Sequence="26" Comments="" Status="0" Hard-coded="False" Audio="False">
    + <Versions>
    + <Item ID="10898" Collection-ID="" Collection-dbtable="" Collection-dbtype="" Analytic-ID="" SMQ620_" Type="21" Lookup-field="" SAS-type="1" SAS-length="0" SAS-format="" SAS-label="" Display-type-params="" CAPI-instructions=""/>
    + <Targets>
  </Node>
- <Node ID="10682" Name="SMQ620" Path="NHANES-05\MI-05\MI-ACASI\SMQ\SMQ620" Type="4" Description="" Index="1" Sequence="32" Comments="" Status="0" Hard-coded="True" Audio="False">
  + <Versions>
  + <Item ID="10604" Collection-ID="" Collection-dbtable="" Collection-dbtype="smailint" Analytic-ID="" SMQ620" Type="5" Lookup-field="" SAS-type="0" SAS-length="1" SAS-format="1." SAS-label="" Display-type-params="" CAPI-instructions=""/>
  + <Targets>
  </Node>
- <Node ID="10684" Name="SMQ630" Path="NHANES-05\MI-05\MI-ACASI\SMQ\SMQ630" Type="4" Description="" Index="2" Sequence="34" Comments="" Status="0" Hard-coded="True" Audio="False">
  + <Versions>
  + <Item ID="10606" Collection-ID="" Collection-dbtable="" Collection-dbtype="smailint" Analytic-ID="" SMQ630" Type="22" Lookup-field="" SAS-type="0" SAS-length="2" SAS-format="2." SAS-label="" Display-type-params="" CAPI-instructions=""/>
  + <Targets>
  </Node>
- <Node ID="10685" Name="SMQ640" Path="NHANES-05\MI-05\MI-ACASI\SMQ\SMQ640" Type="4" Description="" Index="3" Sequence="37" Comments="" Status="0" Hard-coded="False" Audio="False">
  + <Versions>
  + <Item ID="10607" Collection-ID="" Collection-dbtable="" Collection-dbtype="smailint" Analytic-ID="" SMQ640" Type="22" Lookup-field="" SAS-type="0" SAS-length="2" SAS-format="2." SAS-label="" Display-type-params="" CAPI INSTRUCTION: ACCEPTABLE VALUES: 0-30, Refused, Don't Know If SMQ640 > 30 then ERROR ("Your response cannot exceed 30 days. Please try again.") |____|____| ENTER NUMBER OF DAYS REFUSED 77 DON'T KNOW 99">
  + <Targets>
<Node ID="10682" Name="SMQ620" Path="NHANES-05\MI-05\MI-ACASI\SMQ\SMQ620" Type="4" Description="" Index="1" Sequence="32" Comments="" Status="0" Hard-coded="True" Audio="False">
+ <Versions>
- <Item ID="10604" Collection-ID="SMQ620" Collection-dbtable="" Collection-dbtype="smallint" Analytic-ID="SMQ620" Type="5" Lookup-field="" SAS-type="0" SAS-length="1" SAS-format="1." SAS-label="Ever tried cigarette smoking" Display-type="0" Display-type-params="" CAPI-instructions="">
- <Codes>
  - <Code ID="26666" Item-collection-ID="SMQ620" Type="1" Name="" Index="0" Selected="False" Description="" Value="1" Fill-value="" Skip-node-ID="0" Skip-node-name="" EOS="False" Code-preset="">
    <Label Language="English">Yes</Label>
    <Label Language="Spanish">Sí</Label>
  - <Sounds Language="English">
    <Sound Path="YES.wav" />
  </Sounds>
  - <Sounds Language="Spanish">
    <Sound Path="YES.wav" />
  </Sounds>
  - <Ranges>
    <Range ID="0" Name="Soft Edit" Description="Soft Edit" Type="0" From="" To="" />
    <Range ID="0" Name="Hard Edit" Description="Hard Edit" Type="1" From="" To="" />
  </Ranges>
</Code>
- <Code ID="26667" Item-collection-ID="SMQ620" Type="1" Name="" Index="1" Selected="False" Description="" Value="2" Fill-value="" Skip-node-ID="10983" Skip-node-name="SMQ680" EOS="False" Code-preset="">
  <Label Language="English">No</Label>
  <Label Language="Spanish">No</Label>
  - <Sounds Language="English">
    <Sound Path="NO.wav" />
  </Sounds>
  - <Sounds Language="Spanish">
    <Sound Path="NO.wav" />
  </Sounds>
  - <Ranges>
</Code>
</Item>
</Versions>
</Node>
<Pattern ID="169" Text="{su(SE) Número de Seguro Social/el Número de Seguro Social de SP}">
  <Substitution Type="5" Text="el Número de Seguro Social de FULL_NAME"/>
  <Substitution Type="1" Text="su Número de Seguro Social"/>
</Pattern>

- <Pattern ID="257" Text="{él(MPR)/ella(FPR)}">
  <Substitution Type="7" Text="ella"/>
  <Substitution Type="6" Text="él"/>
</Pattern>

- <Pattern ID="168" Text="{su(SE) raza/la raza de SP}">
  <Substitution Type="5" Text="la raza de FULL_NAME"/>
  <Substitution Type="1" Text="su raza"/>
</Pattern>

</LanguagePatterns>

- <LanguagePatterns Language="English">
  - <Pattern ID="25" Text="{carry/carries}"/>
    <Substitution Type="5" Text="carries"/>
    <Substitution Type="1" Text="carry"/>
</Pattern>

  - <Pattern ID="6" Text="{his/her/your}"/>
    <Substitution Type="7" Text="her"/>
    <Substitution Type="6" Text="his"/>
    <Substitution Type="1" Text="your"/>
</Pattern>

  - <Pattern ID="343" Text="{you/your child}"/>
    <Substitution Type="5" Text="your child"/>
    <Substitution Type="1" Text="you"/>
</Pattern>

  - <Pattern ID="149" Text="{you were}"/>
    <Substitution Type="1" Text="you were"/>
</Pattern>

  - <Pattern ID="269" Text="{you have/s/he/SP has}"/>
    <Substitution Type="5" Text="FULL_NAME has"/>
</Pattern>
<?xml version="1.0" standalone="no" ?>
<!-- Common sounds: Created on 1/17/2006 -->
<CommonSounds>
  <LanguageSounds Language="English">
    <Sound Type="1" Path="Retry.wav" />
    <Sound Type="2" Path="Next.wav" />
    <Sound Type="3" Path="WhyNot.wav" />
    <Sound Type="4" Path="REFUSAL.wav" />
    <Sound Type="5" Path="DontKnow.wav" />
    <Sound Type="6" Path="InactivitySound_EN.wav" />
  </LanguageSounds>
  <LanguageSounds Language="Spanish">
    <Sound Type="1" Path="Retry.wav" />
    <Sound Type="2" Path="Next.wav" />
    <Sound Type="3" Path="WhyNot.wav" />
    <Sound Type="4" Path="REFUSAL.wav" />
    <Sound Type="5" Path="DontKnow.wav" />
    <Sound Type="6" Path="InactivitySound_SP.wav" />
  </LanguageSounds>
</CommonSounds>
<table>
<thead>
<tr>
<th>lookup fld</th>
<th>OTC</th>
<th>lookup key</th>
<th>KAVA KAVA ADVANCEDMULA</th>
<th>Kava (Piper methysticum)</th>
<th>Kava (Piper methysticum)</th>
<th>Kava (Piper methysticum)</th>
</tr>
</thead>
<tbody>
<tr>
<td>lookup fld</td>
<td>OTC</td>
<td>lookup key</td>
<td>KAVA KAVA</td>
<td>Kava (Piper methysticum)</td>
<td>Kava (Piper methysticum)</td>
<td>Kava (Piper methysticum)</td>
</tr>
<tr>
<td>lookup fld</td>
<td>OTC</td>
<td>lookup key</td>
<td>KAVA KAVA ADVANCEDMULA</td>
<td>Kava (Piper methysticum)</td>
<td>Kava (Piper methysticum)</td>
<td>Kava (Piper methysticum)</td>
</tr>
</tbody>
</table>
Metadata in Action
Now I will ask about TV watching. Over the past 30 days, on a typical day how much time altogether did you spend sitting and watching TV or videos? Would you say...

- Less than 1 hour,
- 1 hour,
- 2 hours,
- 3 hours,
- 4 hours,
- 5 hours or more, or
- None (you do/SD does) not use a computer outside of work?
Which of these are reasons you stopped being covered by or do not have health insurance?

HAND CARD HIQ3. CODE ALL THAT APPLY.

- Person in family with health insurance lost job or changed employers
- Got divorced or separated/death of spouse or parent
- Became ineligible because of age/left school
- Employer does not offer coverage/or not eligible for coverage
- Cost is too high
- Insurance company refused coverage
- Medicaid/medical plan stopped after pregnancy
- Lost medicaid/medical plan because of new job or increase in income
- Lost medicaid (other)
- Other (specify) Out of country
Please collect all samples

<table>
<thead>
<tr>
<th>Name</th>
<th>Volume</th>
<th>Bar Code</th>
<th>Collected</th>
<th>Timestamp</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gray Top</td>
<td>2 ml</td>
<td>11223333</td>
<td>Fully</td>
<td>1/16/2006 1:01:45 PM</td>
<td></td>
</tr>
<tr>
<td>Gold Top</td>
<td>4 ml</td>
<td>11223334</td>
<td>Partially</td>
<td>1/16/2006 1:01:47 PM</td>
<td>Equipment Failure</td>
</tr>
<tr>
<td>Lavender Top</td>
<td>3 ml</td>
<td>11223335</td>
<td>Fully</td>
<td>1/16/2006 1:01:48 PM</td>
<td></td>
</tr>
<tr>
<td>Lavender Top</td>
<td>3 ml</td>
<td>11223336</td>
<td>Partially</td>
<td>1/16/2006 1:01:50 PM</td>
<td></td>
</tr>
<tr>
<td>Red Top</td>
<td>10 ml</td>
<td>11223337</td>
<td>None</td>
<td>1/16/2006 1:01:51 PM</td>
<td>Equipment Failure</td>
</tr>
<tr>
<td>Red Top</td>
<td>10 ml</td>
<td>11223338</td>
<td>Partially</td>
<td>1/16/2006 1:01:55 PM</td>
<td>Equipment Failure</td>
</tr>
</tbody>
</table>

Overall Comment
Oxidant Badge

Please check the Expiration Date for the Air Collection Device. If the Expiration Date is not within the acceptable limits, obtain a backup device. If a backup device is not available, do not perform sample collection until an acceptable device is obtained. Record Comments for that device below.

<table>
<thead>
<tr>
<th>Device Name: Oxidant Badge</th>
<th>Started: 1/16/2006 1:33:57 PM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Device Barcode: 12345</td>
<td>Ended: Press to Stop</td>
</tr>
<tr>
<td>Device Expiration Date: 1/1/2010</td>
<td>Secure According to SOP?: ✓</td>
</tr>
<tr>
<td>Location: Master Bedroom</td>
<td>Evidence of Damage?: ✓</td>
</tr>
<tr>
<td>Location in Room: Above Bed</td>
<td>Comments:</td>
</tr>
<tr>
<td>Distance off the Floor (ft): 1.1</td>
<td>Period: 24 Hour</td>
</tr>
<tr>
<td>Confounding Factors: Open Windows</td>
<td>Elapsed Time: 01 min 41 sec</td>
</tr>
<tr>
<td>Container Barcode: 987654</td>
<td>Remaining Time: 23 hour 56 min 18 sec</td>
</tr>
<tr>
<td>Comment:</td>
<td></td>
</tr>
</tbody>
</table>
Water Sample Collection - VOA

VOA Vial 1

- Container Name: VOA Vial 1
- Container ID: 123456
- Location: Bathroom
- Water Filter Used?: Yes
- Filter On When Collected?: Yes
- Preservatives Added?: Yes
- Date/Time of Collection: [field]
- Collected According to SOP?: Yes
- Observations About Water: Clear
- Comments: 

VOA Vial 2

- Container Name: VOA Vial 2
- Container ID: 123457
- Location: Bathroom
- Water Filter Used?: Yes
- Filter On When Collected?: Yes
- Preservatives Added?: Yes
- Date/Time of Collection: [field]
- Collected According to SOP?: Yes
- Observations About Water: Milky
- Comments: Device Relocated

VOA Vial 3

- Container Name: [field]
- Container ID: [field]
- Location: [field]
- Water Filter Used?: [field]
- Filter On When Collected?: [field]
- Preservatives Added?: [field]
- Date/Time of Collection: [field]
- Collected According to SOP?: [field]
- Observations About Water: [field]
- Comments: [field]
ACASI: Yes/No

Here is an example of one type of question: ‘Are you tired?’

Please select ...

Yes  No

Back  Clear  Next
ACASI: Numeric Pad

Another type of question requires a number response. Answer by pressing the number buttons on the keypad. 'How many hours did you sleep last night?'

Please enter a number.
How old were you when you smoked a whole cigarette for the **first** time?

Error

Your response is older than your recorded age.

Please press the "Back" button, press "Clear" and try again.
During the past 30 days, on the days that you smoked, which brand of cigarettes did you usually smoke?

Please select one of the following choices:

- Marlboro
- Camel
- Newport
- Kool
- Winston
- Benson and Hedges
- Salem
- Some other brand
Technology Used

- .net 1.1, .net 2.0, C#
- MS SQL Server 2000, 2005
- XML
- DirectX
What Next?

• Integration with other metadata sources
• Complying with metadata standards
• Shared repository
Summary

• Metadata is heart of CHITA software suite
• Metadata is cleanly separated from software and maintained by visual editor
• Metadata stored in relational database and exported/imported to/from XML
• Master repository can be shared between surveys and organizations
• Metadata used for data collection, delivery and analysis
• Metadata used for CAPI, ACASI, Exams and other components