

Improving NHANES Data Documentation Processes

FEDCASIC March 21, 2013

Presenter: Jennifer Dostal, CDC/NCHS

Co-authors: Shannon Corcoran, Ed Stammerjohn, Tim Tilert,
Jane Zhang, CDC/NCHS



**Centers for Disease Control and Prevention
National Center for Health Statistics**



Objectives

Describe:

- the challenges NHANES faced, and
- how it resolved many of them with a custom software application.



Photo credit : [sector7gproductions/Photobucket](#)

Background

- National Health and Nutrition Examination Survey
- Collects health data from a nationally representative sample of approximately 10,000 participants every two years
- Releases 100+ datasets every two years via its website



Background – Publishing Process

- Started in 2002
- Monthly web publication of SAS datasets and documentation
- Data documentation, variable lists, and webpages were created/edited individually
- Only codebook portion of documentation was stored in central database
- ~7-10 business days for web publication

Overview of Challenges

- Increasing maintenance
- Section 508 compliance
- Evolving format of documentation
- Reliance on hand-editing
- Loose adherence to storage guidelines



Photo credit: purplemoon98/Photobucket

Challenges

Increasing amount of maintenance

- Every two years NHANES adds more webpages
- Reconciliation of different metadata formats
- Reconciliation of multiple survey cycles
- All changes to the layout of the webpages were made by hand



Challenges

Increasing requirements of 508 compliance

- Section 508 requirements have increased
- Requires every PDF publisher to have specific knowledge
- Most changes required are repetitive
- Compliance a moving target
- Correcting errors could be difficult



Challenges

Evolving format of documentation

- Documentation format evolution
- 3 documentation files → 2 documentation files → 1 documentation file
- Mixture of formats didn't make sense from the outside

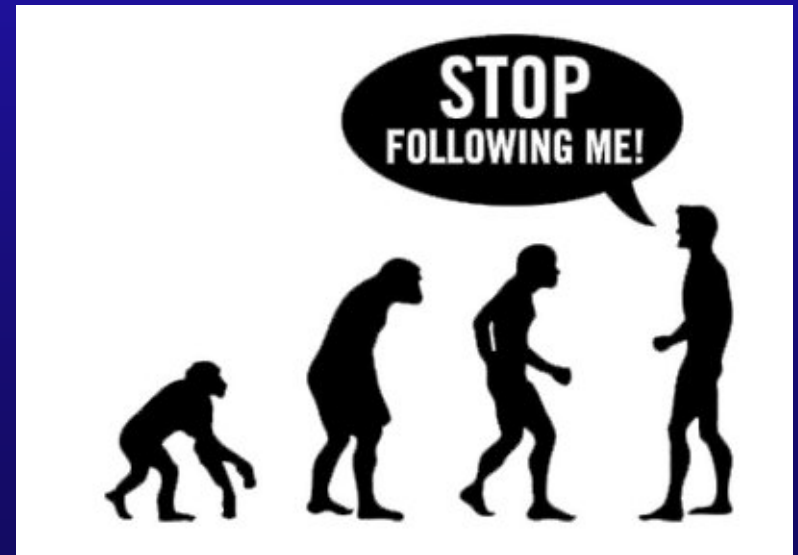


Photo credit: serendipity35.net

Challenges

Reliance on hand-editing webpages

- Detailed instructions
- All changes had to be thoroughly thought out and account for cascading changes
- Required detailed knowledge of website file structure and web publishing processes

Challenge

Loose adherence to storage guidelines

- Personal responsibility
- Individual preferences



Ideal System Wishlist

- Section 508 compliant data documentation
- Work for all documentation
- Centrally manage metadata
- Fewer steps to web publication
- Template-based products



Photo credit: theblackcloset/Photobucket

Meeting the Challenges

- Project X is NHANES data documentation management system
- Produces 508-compliant HTML documentation with a minimum of user input
- Databases store final metadata
- Streamlines publication process

What is Project X?

- NHANES data documentation management system
- Stores and publishes data documentation and related files
- Custom built application using SQL Server and ASP.Net. Front-end also uses Telerik controls.
- Backed by two databases

Project X

- Home
- Edit
- Documentation
- 2011-2012
- 2009-2010
- 2007-2008
- 2005-2006
- 2003-2004
- 2001-2002
- 1999-2000
- PHANES III
- PHES I
- and Publish
- Laborators
- 2011-2012
- 2009-2010
- 2007-2008
- 2005-2006
- 2003-2004
- 2001-2002
- 1999-2000
- PHANES III
- PHES I
- and Publish
- ic
- 2011-2012
- 2009-2010
- 2007-2008
- 2005-2006
- 2003-2004
- 2001-2002
- 1999-2000
- PHANES III
- PHES I
- iew What's
- ate Web Lists
- ate Variable
- ate What's
- List

Data Documentation for 2011-2012

The documentation for this cycle is grouped into Components. Click one of the blue tabs at the top to display the list of documentation for a different component.

To **add a new document**, click the Add New Record heading or green button at the top or bottom of the list. This opens the Insert form. Enter the information about the document, and click Insert at the bottom of the form to submit it to the database. Or, click Cancel to cancel your changes.

To **edit the cover information** for the documentation, click the Pencil icon to the left of the documentation title. This will open the form for editing and show you all the cover information for the document. Click the Update link at the bottom of the form to submit your changes to the database. Or, click Cancel to cancel your changes.

To **edit the Section, References, or Appendices contents**, click the appropriate hyperlink on the right of the title.

To **preview** the documentation as a webpage, click the Preview hyperlink.

To **copy** the cover information and contents of the documentation into a new document, click the Copy hyperlink.

To **sort the list**, click the DocTitle and DocID headings to sort them in ascending order, or click twice to sort them in descending order.

Demographics Examination **Laboratory** Questionnaire Dietary

Label

| + Add new record | | | | | | | | Refresh |
|------------------|---|----------|--------------------------|----------------------------|----------------------------|-------------------------|----------------------|---------|
| DocID | DocTitle | Abbr | Sections | References | Appendices | Preview | Copy | |
| | 1137 Blood Lead, Cadmium, Total Mercury, Selenium, and Manganese | PbCd_G | Sections | References | Appendices | Preview | Copy | |
| | 1140 Blood mercury: inorganic, ethyl and methyl | IHgEM | Sections | References | Appendices | Preview | Copy | |
| | 1141 Complete Blood Count with 5-Part Differential in Whole Blood | CBC_G | Sections | References | Appendices | Preview | Copy | |
| | 1156 Environmental Pesticides | PP_G | Sections | References | Appendices | Preview | Copy | |
| | 1133 Hepatitis A | HEPA_G | Sections | References | Appendices | Preview | Copy | |
| | 1150 Hepatitis B surface antibody | HEPB_S_G | Sections | References | Appendices | Preview | Copy | |

| | | | | | | |
|--|------|---|----------|--------------------------|----------------------------|----------------------------|
| | 1143 | Human Papillomavirus (HPV) in Oral Rinse | ORHPV_G | Sections | References | Appendices |
| | 1144 | Human Papillomavirus (HPV) in Oral Rinse, Ages 14-17 | OHPV_G_R | Sections | References | Appendices |
| | 1129 | Serum Cotinine | COT_G | Sections | References | Appendices |
| | 1145 | Serum Cotinine and Urinary Total NNAL | COTNAL_G | Sections | References | Appendices |
| | 1139 | Standard Biochemistry Profile | BIOPRO_G | Sections | References | Appendices |
| | 1157 | Thyroid Profile | THYROD_G | Sections | References | Appendices |
| | 1142 | Tissue transglutaminase assay (IgA-TTG) and the IgA endomyseal antibody assay (IgA EMA) | TGEMA_G | Sections | References | Appendices |
| | 1130 | Total Cholesterol | TCHOL_F | Sections | References | Appendices |

Edit details for Total Cholesterol

| | | | | |
|-------------------------------|--|--|-------------------------------|--------------------------|
| DocTitle: | <input type="text" value="Total Cholesterol"/> | Grouping <i>(Only use with Non-public, Surplus Sera and Second Day Exam/Lab datasets.):</i> | ApprovedforCollaborators: | <input type="checkbox"/> |
| BeginYear: | <input type="text" value="2011"/> | | ApprovedforCollaboratorsDate: | <input type="text"/> |
| EndYear: | <input type="text" value="2012"/> | Main | ApprovedforCollaboratorsBy: | <input type="text"/> |
| Abbr: | <input type="text" value="TCHOL_F"/> | | PublishedforCollaborators: | <input type="checkbox"/> |
| Access Constraints: | <input type="text" value="None"/> | | ApprovedforPublic: | <input type="checkbox"/> |
| Use Constraints: | <input type="text" value="None"/> | XPTFileSize: | ApprovedforPublicDate: | <input type="text"/> |
| Component: | <input type="text" value="Laboratory"/> | DataFileType (XPT, EXE, TXT, ZIP): | ApprovedforPublicBy: | <input type="text"/> |
| Format of Data Documentation: | <input type="text" value="Laboratory"/> | SpecialNote: | PublishedforPublic: | <input type="checkbox"/> |
| Subject: | <input type="text" value="Total Cholesterol"/> | Uses Subsample Weight: | | |
| FirstPublished: | <input type="text" value="9/30/2013 12:00:00 AM"/> | | | |
| LastRevised: | <input type="text"/> | | | |

Project X

- Home
- Add/Edit
- Documentation
 - 2011-2012
 - 2009-2010
 - 2007-2008
 - 2005-2006
 - 2003-2004
 - 2001-2002
 - 1999-2000
 - NHANES III
 - NHES I
- View and Publish
- Collaborators
 - 2011-2012
 - 2009-2010
 - 2007-2008
 - 2005-2006
 - 2003-2004
 - 2001-2002
 - 1999-2000
 - NHANES III
 - NHES I
- View and Publish
- Public
 - 2011-2012
 - 2009-2010
 - 2007-2008
 - 2005-2006
 - 2003-2004
 - 2001-2002
 - 1999-2000
 - NHANES III
 - NHES I
- Preview What's New
- Create Web Lists

Move to: [Sections](#) | [References](#) | [Appendices](#) | [Preview this Documentation](#)

Total Cholesterol (TCHOL_F)

+ Add new record

| | Edit | Section Title | Section Contents |
|------------------------|------|-----------------------|---|
| Delete | | Component Description | The goals of this component are: 1) to monitor the prevalence and trends in major cardiovascular risk factors in the U.S.; and 2) to evaluate prevention and treatment programs targeting cardiovascular disease in the U.S. The main element of the cardiovascular disease laboratory component in NHANES is blood lipids. Cardiovascular disease is the leading cause of death in the United States. The data will be used to monitor the status of hyperlipidemia and the success of the National Cholesterol Education Program. |

| | | |
|------------------------|-----------------|--|
| Delete | Eligible Sample | Participants aged 6 years and older were tested. |
|------------------------|-----------------|--|

Section Title: Eligible Sample

Rich text editor toolbar with icons for Bold, Italic, Underline, Text Color, Background Color, Bulleted List, Numbered List, Undo, Redo, Print, and ABC.

Rich text editor toolbar with icons for Insert Image, Insert Link, Insert Table, and Help.

Participants aged 6 years and older were tested.

Section Contents:

Project X

- Home
- Add/Edit
- Documentation
 - 2011-2012
 - 2009-2010
 - 2007-2008
 - 2005-2006
 - 2003-2004
 - 2001-2002
 - 1999-2000
 - NHANES III
 - NHES I
- View and Publish Collaborators
 - 2011-2012
 - 2009-2010
 - 2007-2008
 - 2005-2006
 - 2003-2004
 - 2001-2002
 - 1999-2000
 - NHANES III
 - NHES I
- View and Publish Public
 - 2011-2012
 - 2009-2010
 - 2007-2008
 - 2005-2006
 - 2003-2004
 - 2001-2002
 - 1999-2000
 - NHANES III
 - NHES I
- Preview What's New
- Create Web Lists

Publish Public Data Documentation for 2009-2010

Documentation that **has not been published for public release** is highlighted in yellow. Click the "View Doc" column heading to view the documentation.

Click the "Publish" link to publish or republish documentation as public documentation. After documentation has been published, click the "View" hyperlink to view it.

| Demographics | | Examination | | Laboratory | | Questionnaire | | Dietary | |
|--|----------|-------------|------------|------------|-----------------|---------------|-------------------------|----------------------|-----|
| DocTitle | Abbr | Component | Begin Year | End Year | First Published | Last Revised | Publish Doc | View Doc | Acc |
| Apolipoprotein B | ApoB_F | Laboratory | 2009 | 2010 | June, 2012 | | Publish | View | Non |
| Blood Cadmium and Lead | PbCd_F | Laboratory | 2009 | 2010 | September, 2011 | | Publish | View | Non |
| Blood Total Mercury and Blood Inorganic Mercury | THgIHg_F | Laboratory | 2009 | 2010 | September, 2011 | | Publish | View | Non |
| Complete Blood Count with 5-Part Differential in Whole Blood | CBC_F | Laboratory | 2009 | 2010 | September, 2011 | January, 2012 | Publish | View | Non |
| C-Reactive Protein | CRP_F | Laboratory | 2009 | 2010 | September, 2011 | | Publish | View | Non |
| Environmental Pesticides | PP_F | Laboratory | 2009 | 2010 | October, 2011 | | Publish | View | Non |
| Environmental Phenols | EPH_F | Laboratory | 2009 | 2010 | October, 2011 | | Publish | View | Non |
| Epstein-Barr virus (VCA IgG) | SSEBV_F | Laboratory | 2009 | 2010 | June, 2012 | | Publish | View | Non |
| Fasting Questionnaire | FASTQX_F | Laboratory | 2009 | 2010 | September, 2011 | | Publish | View | Non |

National Health and Nutrition Examination Survey

2009 - 2010 Data Documentation, Codebook, and Frequencies

Total Cholesterol (TCHOL_F)

Data File: TCHOL_F.xpt

First Published: September, 2011

Last Revised: NA

Locator Record

Title: Total Cholesterol (TCHOL_F)
Contact Number: 1-866-441-NCHS
Years of Content: 2009 - 2010
First Published: September, 2011
Revised: NA
Access Constraints: None
Use Constraints: None
Geographic Coverage: National
Subject:Total Cholesterol
Record Source: NHANES 2009 - 2010
Survey Methodology: NHANES 2009 - 2010 is a stratified multistage probability sample of the civilian non-institutionalized population of the U.S.
Medium: NHANES Web site; SAS transport files

Table of Contents

- [Component Description](#)
- [Eligible Sample](#)
- [Description of Laboratory Methodology](#)
- [Laboratory Quality Assurance and Monitoring](#)
- [Data Processing and Editing](#)
- [Analytic Notes](#)
- [References](#)
- [Codebook](#)
 - [SEQN - Respondent sequence number](#)
 - [LBXTC - Total Cholesterol \(mg/dL\)](#)
 - [LBDTCSI - Total Cholesterol \(mmol/L\)](#)

Browser address bar: http://adp-dhanes/NCHS/NHANES/nhanes2009-2010/TCHOL_F.htm#SEQN

Search bar: Find: Previous Next Options

Navigation icons: Home, RSS, Print, Page, Safety

Bookmarks: Favorites, GovDelivery » Bulletins, ASP.NET Community ..., TinyURL!, Get more Add-ons

Open tabs: CDC NHANES - NHANES 2009-..., NHANES 2009 - 2010: T..., Bulletins Report

- Roche/Hitachi System Application Sheet for HDL-C plus 3rd generation, 2007.

Codebook and Frequencies

SEQN - Respondent sequence number

Variable Name: SEQN
SAS Label: Respondent sequence number
English Text: Respondent sequence number.
Target Gender: Both males and females
Target Age: 6 YEARS - 150 YEARS

LBXTC - Total Cholesterol (mg/dL)

Variable Name: LBXTC
SAS Label: Total Cholesterol (mg/dL)
English Text: Total Cholesterol (mg/dL)
Target Gender: Both males and females
Target Age: 6 YEARS - 150 YEARS

| Code or Value | Value Description | Count | Cumulative | Skip to Item |
|---------------|-------------------|-------|------------|--------------|
| 66 to 528 | Range of Values | 7846 | 7846 | |
| . | Missing | 745 | 8591 | |

Table of Contents

- [Component Description](#)
- [Eligible Sample](#)
- [Description of Laboratory Methodology](#)
- [Laboratory Quality Assurance and Monitoring](#)
- [Data Processing and Editing](#)
- [Analytic Notes](#)
- [References](#)
- [Codebook](#)
 - [SEQN - Respondent sequence number](#)
 - [LBXTC - Total Cholesterol \(mg/dL\)](#)
 - [LBDTCSI - Total Cholesterol \(mmol/L\)](#)

Meeting the Challenges

- Project X is NHANES data documentation management system
- **Produces 508-compliant HTML documentation with a minimum of user input**
- Databases store final metadata

Producing 508 compliant documentation

- HTML files are easier to make compliant.
- Content requires little tweaking.
- Supports the implementation of library-wide changes in the template.

```
<html>  
<body>  
I love html  
</body>  
</html>
```

Meeting the Challenges

- Project X is NHANES data documentation management system
- Produces 508-compliant HTML documentation with a minimum of user input
- **Databases store final metadata**

Databases

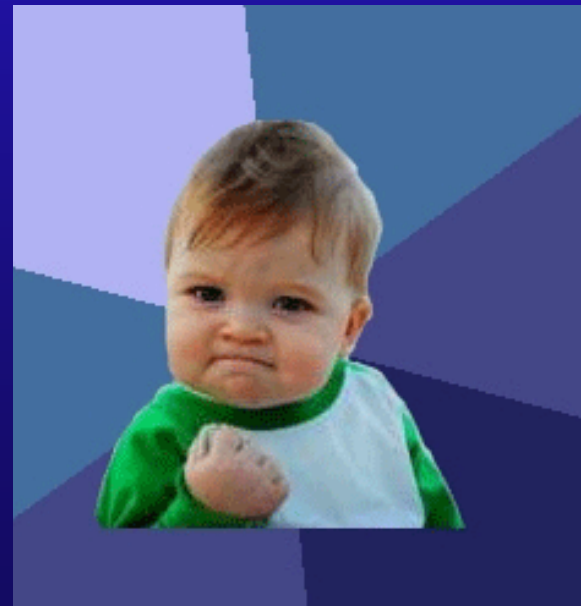
- Two databases
- Final storage of all metadata
- Promotes creation of new user tools
- Easily present the same metadata in different formats



Trademark of Microsoft

Additional Benefits

- Faster publishing of documentation, related metadata, and web pages
- Developed basic variable search for data users
- No additional work was needed to support an additional survey



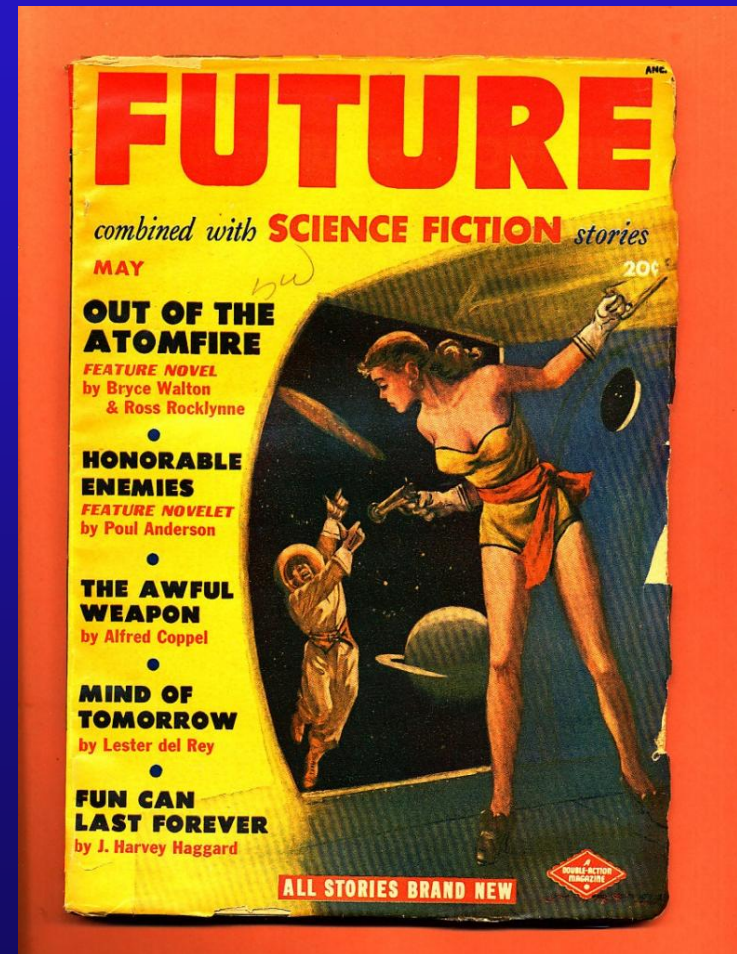
Trade-offs

- Build it
- Timing
- Moving contents
Simplification and mass production
- Address concerns of other stakeholders
 - web publishing group
 - external data users



Possible Future Development

- Expanded variable search
- Datasets in multiple formats
- Direct access to related files
- Basic metadata at a glance
- RSS feed of new and updated files
- Internal dashboard



Photocredit: smilingdawg/Photobucket

End of File

- I'm done.
- Let's talk!

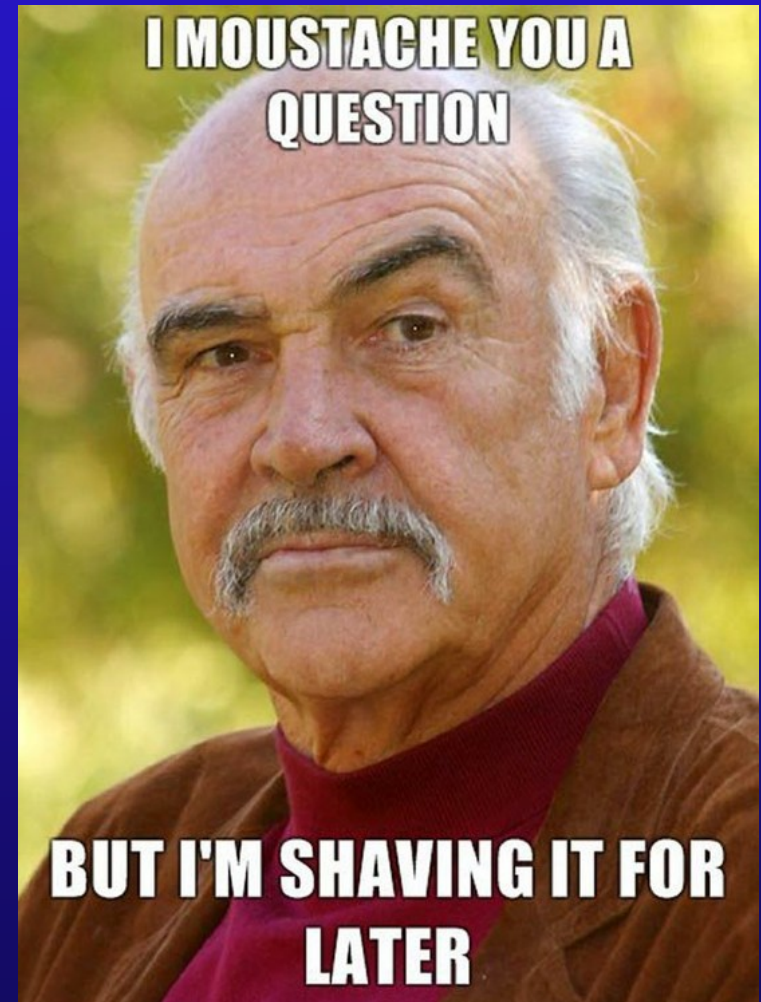


Photo Credit: George Takei/Tumblr