

Discover the Power of DDI Metadata

Monday March 31, 2014
Workshop
North American DDI User Conference
<naddi>

Jane Fry
Carleton University



Schedule

- Introductions
- Brief background of DDI
- Exercise 1
- Getting started
- Examples
- Integration into a data lifecycle workflow
- Exercise 2
- Wrap-up



Brief Background of DDI

- What is DDI?
 - Data Documentation Initiative
 - <http://www.ddi-alliance.org/>
- An international specification
 - Not yet a formal ISO standard
 - Goal
- Formats documentation for a social science data file
 - More useful than a word or text file
- Supports the entire research data lifecycle

Brief Background (cont'd)

- Creates a standard format
 - Used to mark up codebooks
 - Meaningful and consistent
 - Metadata is both human and machine readable
- Gives codebook level details such as
 - dataset contents, variable labels, summary Statistics and frequencies
 - Also question text for each variable
 - *as long as it was entered when marking up the document*

Brief Background (cont'd)

- DDI website (<http://www.ddialliance.org/what>)
 - Excellent resource
 - FAQ
 - Tools
 - Markup Examples
 - Metadata Resources



Brief Background (cont'd)

- Facilitates the creation of metadata
- Expressed in XML
 - XML Schema
 - A way of tagging text for meaning, not appearance
 - Older version of DDI used DTD (Document type definition)
 - Defines
 - Which tags are available
 - The order the tags will appear in a document
 - Whether the tags are required or optional
 - Whether the tags are repeatable or not

Example of DDI Compliant Codebook

```
<?xml version='1.0' encoding='UTF-8'?>
<codeBook version="1.2.2" ID="ctums_82Moo2oXCB_E_2004_ann_person-file">
  <docDscr>
    <citation>
      <titlStmt>
        <titl>
          Canadian Tobacco Use Monitoring Survey, 2004: Annual, Person File
        </titl>
        <subTitl>
          Annual, Person File
        </subTitl>
        <altTitl>
          CTUMS 2004: Annual, Person File
        </altTitl>
        <parTitl>
          Enquete de surveillance de l'usage du tabac au Canada, 2004: Annuel, Personne
        </parTitl>
        <IDNo>
          ctums_82Moo2oXCB_E_2004_ann_person-file
        </IDNo>
      </titlStmt>
      <rspStmt>
        <AuthEnty affiliation="Queen's University. Social Science Data Centre">
          Cooper, Alex
        </AuthEnty>
      </rspStmt>
    </citation>
  </docDscr>
</codeBook>
```



Brief Background (cont'd)

- Benefits of DDI
 - Interoperability
 - Rich content
 - A single document with multiple purposes
 - On-line subsetting and analysis
 - Increased search capability
 - precision in searching
 - More visibility for the data because it is web-based
 - Research is no longer on just one computer



Brief Background (cont'd)

- Challenges
 - Complexity
 - Level of researcher buy-in
 - Need for tools
 - Changes in workflow
 - Access to metadata



Brief Background (cont'd)

- History
 - Started in 1995
 - The director of ICPSR, Richard Rockwell, constituted a Committee on metadata
 - Quebec City
 - Developed a draft list of codebook elements
 - An international committee



Brief Background (cont'd)

- Norwegian Social Science Data Services
- Harvard University
- American University
- Statistics Canada
- Health Canada
- Bureau of the Census
- University of Michigan
- Bureau of Labor Statistics
- ICPSR
- Yale University
- ESRC Data Archive (UK)
- University of California, Berkeley
- University of Southern Denmark
- The Roper Center
- Zentralarchiv für Empirische Sozialforschung (GESIS)



Brief Background (cont'd)

- Bjorn Henrichsen
- Micah Altman
- Atle Alvheim
- Grant Blank
- Ernie Boyko
- Bill Bradley
- Cavan Capps
- Bill Connett
- Cathryn Dipppo
- Pat Doyle
- Dan Gillman
- Peter Granda
- Ann Green
- Peter Joftis
- Ken Miller
- Tom Piazza
- Karsten Boye Rasmussen
- Richard Rockwell
- Jostein Ryssevik
- Merrill Shanks
- Peter Solenberger
- Rolf Uher
- Mary Vardigan



Brief Background (cont'd)

- 1996: Prepared first DTD (document type definition)
 - Set up markups/specifications
- 1997: Received funding from the NSF
 - Specifications translated to XML
- 1999: Beta testing
 - Reports were submitted
- 2000: DDI Version 1 published
- 2001 : First DDI training
- 2001: Received funding from Health Canada

Brief Background (cont'd)

- 2002: DDI Alliance charter drafted
- 2003
 - DDI 2 published
 - DDI Alliance established
 - First Steering Committee meeting
 - Final meeting of original Committee
- Continues to evolve
- Not a one-time thing

Brief Background (cont'd)

- Now branched into 2 separate development lines or metadata standards
 - DDI Codebook
 - aka DDIC
 - Formerly DDI₂
 - This was the original one
 - DDI Lifecycle
 - aka DDIL
 - formerly DDI₃
 - Supports the research data lifecycle
 - The one new users are learning



Brief Background (cont'd)

- DDI Alliance
 - Self-sustaining
 - Membership organization
 - Members have a voice in the development of DDI
 - Membership documents are all online
 - Charter, bylaws, member forms, ...
 - Publications are all one line
 - Newsletters, annual reports, ...
 - <http://www.ddialliance.org/alliance>



Brief Background (cont'd)

- Who is Using DDI?
 - International
 - CESSDA Data Portal (European quantitative social science datasets)
 - Australian Social Science Data Archive
 - DAMES Project (UK)
 - DataFirst at University of Cape Town
 - Data Liberation Initiative (Statistics Canada)
 - General Social Survey (US)
 - ICPSR (Interuniversity Consortium for Political & Social Research)
 - Israel Social Science Data Center
 - Philippines National Statistics Office
 - Statistics New Zealand
 - Vision of Britain (historical view between 1801 and 2001)
 - World Bank (International Household Survey Network)

Global DDI

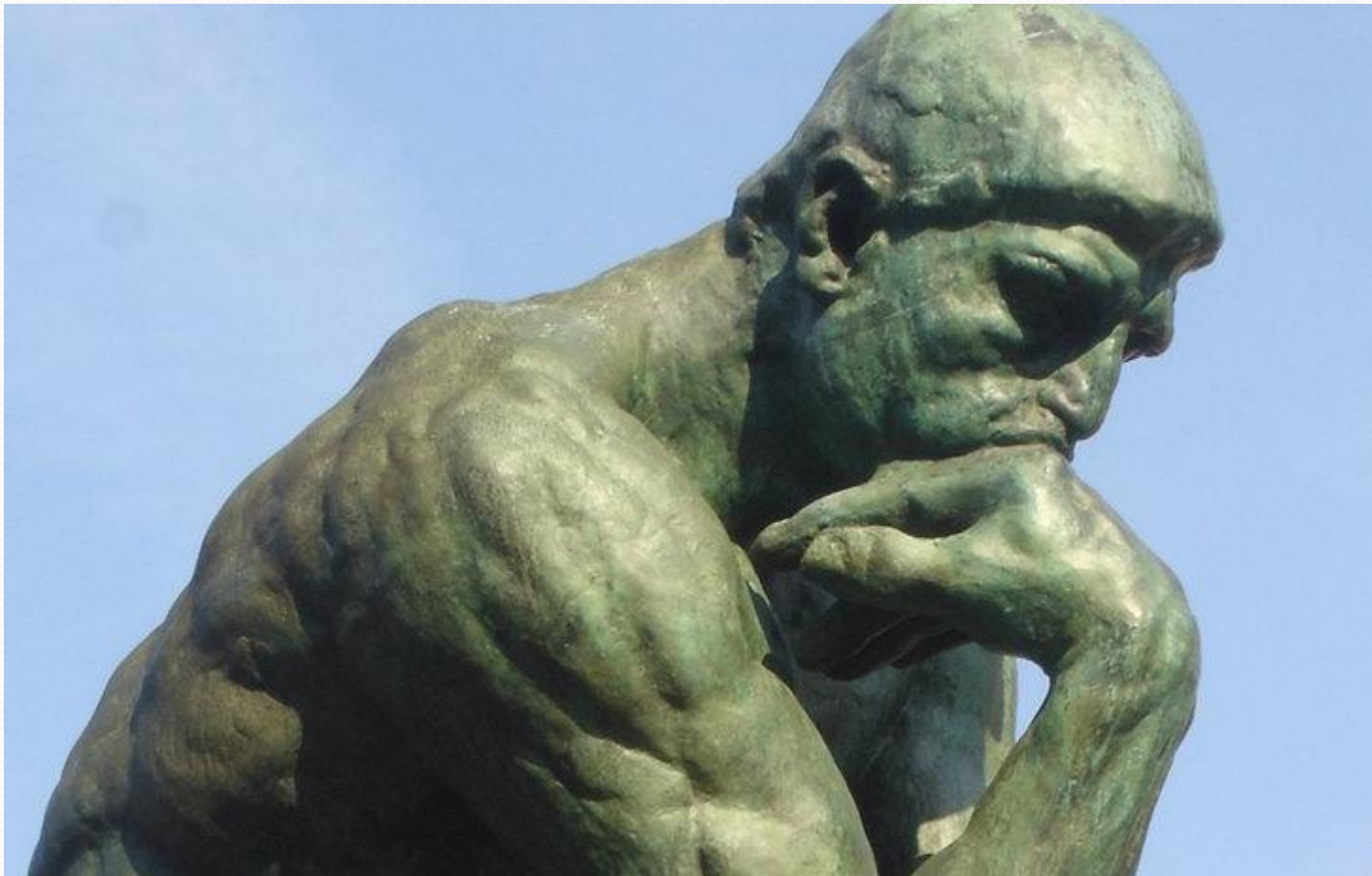




Brief Background (cont'd)

- What Health Projects are using DDI?
 - Altarum Institute (US)
 - Systems research for better health
 - Centers for Medicare/Medicaid Services (CMS) Denominator File Project
 - The Global Fund to Fight AIDS, Tuberculosis and Malaria
 - MIDUS Series
 - Midlife in the US: A National Study of Health and Well-Being
 - The Tromsø Study (Norway)
 - to determine reasons for high mortality rates due to different diseases
 - UNICEF, Child Information
 - Monitoring the Situation of Children and Women (global development data)

Time to think!





Exercise 1

- As a researcher, what metadata do you *absolutely* need?
- How do you want it to *streamline* your research?
- What metadata would you like to have, if it is available, but it is *not integral* to your research?



Getting Started with DDI

- Daunting at first
 - Process is broken down into steps
- Lots of help available
 - DDI Alliance
 - Colleagues
 - Other researchers
- DDI List-serv
- DDI Best Practices
 - Work in progress
 - Feedback always welcome

Getting Started with DDI (cont'd)

- Tools to help you get started
 - <http://www.ddialliance.org/resources/tools>
 - Drop down menu
 - Browse a list
 - Choose the one that will suit your purposes
 - License choices
 - Commercial
 - Freeware
 - GNU Lesser General Public License (LGPL)
 - For different versions of DDI

Getting Started with DDI (cont'd)

- One tool: Nesstar Publisher
 - Norwegian Social Science Data Services
 - Data management program
 - Freeware
 - Data and metadata conversion and editing tools
 - Enhance datasets
 - Combine catalogue and contextual information
 - Merge DDI documents with markup for different sections of the DDI for the same study
 - Merge variable descriptions from SPSS/SAS with DDI

Getting Started with DDI (cont'd)

- Nesstar Webview
 - Metadata
 - Any associated documentation
 - Variable groups
 - Conduct basic analysis
 - Subsetting
 - Crosstabs
 - Bonus

Getting Started with DDI (cont'd)

- Nesstar Webview
 - Downloading
 - Documentation
 - PDF format
 - Export files with study descriptions and question text
 - Data exported in format of choice
 - SPSS, SAS, Stata, ASCII, ...

Getting Started with DDI (cont'd)

- Nesstar drawbacks
 - For advanced statistical analysis -
 - it is best to download the data and use a statistical analysis package
 - Must have access to a server to publish the dataset
 - Not intuitive when starting to markup datasets
 - Not intuitive for first-time user in Webview
- Not a drawback, just a consideration
 - Uses DDI Codebook standard

The screenshot displays the Nesstar Publisher v4.0.9 software interface. The title bar reads "Nesstar Publisher v4.0.9 - cchs-2012.Nesstar - [My Projects]". The menu bar includes "File", "Edit", "Documentation", "Variables", "Data", "Publishing", "Tools", and "Help". The toolbar contains icons for file operations and a language dropdown set to "English".

The interface is divided into three main sections:

- Projects:** A tree view on the left shows the project structure:
 - My Projects
 - cchs-2011-2012
 - Document Description
 - Study Description
 - Other Study Materials
 - Datasets
 - Variable Groups
 - Other Materials
 - External Resources

- File Information:** A central pane displaying metadata for the selected project:
- File Location:** <H:\LDC\CCHS - Cdn Community Health Survey\2011-2012\cchs-2011-2012.Nesstar>
- File Size:** 105.14 MB
- Last Modified:** 2013-11-18 12:15
- Title:** Canadian Community Health Survey, 2011-2012: Annual component
- ID number:** cchs-82M0013-E-2011-2012-Annual-component
- Copyright:** Copyright © Statistics Canada, 2013
- Abstract:**

The Canadian Community Health Survey (CCHS) is a cross-sectional survey that collects information related to health status, health care utilization and health determinants for the Canadian population. It surveys a large sample of respondents and is designed to provide reliable estimates at the health region level. In 2007, major changes were made to the CCHS design. Data is now collected on an ongoing basis with annual releases, rather than every two years as was the case prior to 2007. The survey's objectives were also revised and are as follows:

 - support health surveillance programs by providing health data at the national, provincial and intra-provincial levels;
 - provide a single data source for health research on small populations and rare characteristics;
 - timely release of information easily accessible to a diverse community of users; and
 - create a flexible survey instrument that includes a rapid response option to address emerging issues related to the health of the population.

The CCHS data is always collected from persons aged 12 and over living in private dwellings in the 115 health regions covering all provinces and territories. Excluded from the sampling frame are individuals living on Indian Reserves and on Crown Lands, institutional residents, full-time members of the Canadian Forces, and residents of certain remote regions. The CCHS covers approximately 98% of the Canadian population aged 12 and over.

The CCHS produces three types of microdata files: master files; share files; and public use microdata files (PUMF). The characteristics of each of these files are presented in the User Guide. The PUMF is released every two years and contains two years of data.
- Datasets:**
 - **Name:** cchs-2011-2012
 - **Variable Count:** 1207
 - **Case Count:** 124929

Projects:

- Dates of Collection
- Countries
- Geographic Coverage
 - Geographic Unit
 - Unit of Analysis
 - Universe
 - Kind of Data
- Notes
- Time Method
- Data Collectors
- Frequency of Data Collection
- Sampling Procedure
 - Major Deviations from the Sample Design
 - Mode of Data Collection
 - Type of Research Instrument
 - Data Sources
- Characteristics of the Data Collection Situation
 - Actions to Minimize Losses
 - Control Operations
 - Weighting
 - Cleaning Operations
 - Notes
 - Response Rates
- Estimates of Sampling Error
 - Other Forms of Data Appraisal
 - Class or Status of the Study
 - Locations of Data Collection
 - Restrictions
 - Access Authority
- Citation Requirement
- Deposit Requirement
- Access Conditions
- Disclaimer
- Notes
- Other Study Materials
 - Datasets
 - cchs_2012
 - File Description
 - Key Variables & Relations
 - Variables
 - Data Entry
 - Cell Notes
 - Cube Setups
- Variable Groups
- Other Materials
- External Resources

	1 - VERDATE	2 - GEOGRPR	3 - GEODPM	4 - GEODBCH	5 - ADM_PR	6 - ADM_N0	7 - ADM_N1	8 - ADM_N1	9 - ADM_RN	10 - DHHGAG	11 - DHH_SE	12 - DHHGM	13 - DHHGLE	14 - DHHG61	15 - DHHGL1	16 - DHHGLV	17 - DHHGHS	18 - GEN_01
1	20130606	BRITISH COI	NORTH SHO	VANCOUVER	NO	NOT APPLIC	NO	YES	1	20 TO 24 YE	MALE	SINGLE/NEVE	NONE	NONE	NONE	CHILD 2 PAR	3 PERSONS	GOOD
2	20130606	ONTARIO	DIST. ALGON	NOT APPLIC	NO	NOT APPLIC	YES	NOT APPLIC	2	80 YEARS OF	FEMALE	WIDOW/SEP	NONE	NONE	NONE	UNATTACHE	1 PERSON	VERY GOOD
3	20130606	BRITISH COI	OKANAGAN	INTERIOR	NO	IN PERSON	YES	NOT APPLIC	3	30 TO 34 YE	MALE	COMMON-LA	NONE	1 OR MORE	1 OR MORE	PAR SPOUSE	4 PERSONS	VERY GOOD
4	20130606	QUEBEC	RÉG. LANAU	NOT APPLIC	NO	NOT APPLIC	YES	NOT APPLIC	4	40 TO 44 YE	FEMALE	WIDOW/SEP	NONE	1 OR MORE	1 OR MORE	PARENT CHI	3 PERSONS	VERY GOOD
5	20130606	ONTARIO	YORK REGIC	NOT APPLIC	NO	IN PERSON	YES	NOT APPLIC	5	45 TO 49 YE	FEMALE	MARRIED	NONE	NONE	NONE	SPOUSE / PA	2 PERSONS	VERY GOOD
6	20130606	ALBERTA	SOUTH ZONE	NOT APPLIC	NO	NOT APPLIC	YES	NOT APPLIC	6	65 TO 69 YE	MALE	MARRIED	NONE	NONE	NONE	SPOUSE / PA	2 PERSONS	VERY GOOD
7	20130606	SASKATCHE	SASKATOON	NOT APPLIC	NO	NOT APPLIC	YES	NOT APPLIC	7	55 TO 59 YE	FEMALE	MARRIED	NONE	NONE	NONE	PAR SPOUSE	3 PERSONS	GOOD
8	20130606	ONTARIO	PEEL REGIO	NOT APPLIC	NO	IN PERSON	NO	NO	8	35 TO 39 YE	FEMALE	MARRIED	NONE	1 OR MORE	1 OR MORE	PAR SPOUSE	4 PERSONS	EXCELLENT
9	20130606	ONTARIO	PETERB. COI	NOT APPLIC	NO	IN PERSON	YES	NOT APPLIC	9	50 TO 54 YE	FEMALE	MARRIED	NONE	NONE	NONE	SPOUSE / PA	2 PERSONS	GOOD
10	20130606	ONTARIO	CITY OF OTT	NOT APPLIC	NO	NOT APPLIC	YES	NOT APPLIC	10	65 TO 69 YE	FEMALE	COMMON-LA	NONE	NONE	NONE	SPOUSE / PA	2 PERSONS	GOOD
11	20130606	BRITISH COI	SOUTH VANK	VANCOUVER	NO	IN PERSON	NO	NO	11	30 TO 34 YE	MALE	SINGLE/NEVE	NONE	NONE	NONE	UNATTACHE	2 PERSONS	FAIR
12	20130606	ONTARIO	DURHAM RE	NOT APPLIC	NO	NOT APPLIC	YES	NOT APPLIC	12	60 TO 64 YE	FEMALE	MARRIED	NONE	NONE	NONE	SPOUSE / PA	2 PERSONS	VERY GOOD
13	20130606	BRITISH COI	FRASER NOF	FRASER	YES	NOT APPLIC	YES	NOT APPLIC	13	15 TO 17 YE	FEMALE	SINGLE/NEVE	NONE	1 OR MORE	1 OR MORE	CHILD PARE	3 PERSONS	VERY GOOD
14	20130606	SASKATCHE	SUN COUNT	NOT APPLIC	NO	NOT APPLIC	YES	NOT APPLIC	14	55 TO 59 YE	FEMALE	MARRIED	NONE	NONE	NONE	SPOUSE / PA	2 PERSONS	GOOD
15	20130606	QUEBEC	RÉG. CAP. N	NOT APPLIC	NO	NOT APPLIC	YES	NOT APPLIC	15	65 TO 69 YE	FEMALE	SINGLE/NEVE	NONE	NONE	NONE	UNATTACHE	1 PERSON	VERY GOOD
16	20130606	SASKATCHE	PRINCE ALB	NOT APPLIC	NO	NOT APPLIC	NO	NO	16	25 TO 29 YE	FEMALE	COMMON-LA	NONE	NONE	NONE	SPOUSE / PA	2 PERSONS	VERY GOOD
17	20130606	BRITISH COI	NORTH SHO	VANCOUVER	NO	BOTH	YES	NOT APPLIC	17	45 TO 49 YE	FEMALE	COMMON-LA	NONE	NONE	NONE	SPOUSE / PA	2 PERSONS	EXCELLENT
18	20130606	ONTARIO	WATERLOO	NOT APPLIC	NO	NOT APPLIC	YES	NOT APPLIC	18	55 TO 59 YE	FEMALE	WIDOW/SEP	NONE	NONE	NONE	UNATTACHE	1 PERSON	EXCELLENT
19	20130606	ONTARIO	WATERLOO	NOT APPLIC	NO	NOT APPLIC	YES	NOT APPLIC	19	45 TO 49 YE	MALE	MARRIED	NONE	NONE	NONE	PAR SPOUSE	3 PERSONS	VERY GOOD
20	20130606	ALBERTA	SOUTH ZONE	NOT APPLIC	NO	NOT APPLIC	YES	NOT APPLIC	20	80 YEARS OF	MALE	WIDOW/SEP	NONE	NONE	NONE	UNATTACHE	1 PERSON	VERY GOOD
21	20130606	MANITOBA	PRAIRIE M.	NOT APPLIC	NO	IN PERSON	YES	NOT APPLIC	21	70 TO 74 YE	MALE	MARRIED	NONE	NONE	NONE	SPOUSE / PA	2 PERSONS	GOOD
22	20130606	ONTARIO	HURON, PER	NOT APPLIC	NO	NOT APPLIC	NO	NO	22	25 TO 29 YE	MALE	SINGLE/NEVE	NONE	NONE	NONE	UNATTACHE	1 PERSON	GOOD
23	20130606	BRITISH COI	THOMPSON/	INTERIOR	NO	NOT APPLIC	NO	YES	23	15 TO 17 YE	FEMALE	SINGLE/NEVE	NONE	NONE	NONE	CHILD 2 PAR	3 PERSONS	VERY GOOD
24	20130606	ONTARIO	HURON, PER	NOT APPLIC	NO	IN PERSON	YES	NOT APPLIC	24	45 TO 49 YE	MALE	SINGLE/NEVE	NONE	NONE	NONE	UNATTACHE	1 PERSON	VERY GOOD
25	20130606	QUEBEC	RÉG. COTE-D	NOT APPLIC	NO	NOT APPLIC	YES	NOT APPLIC	25	20 TO 24 YE	MALE	SINGLE/NEVE	NONE	NONE	NONE	CHILD 2 PAR	3 PERSONS	VERY GOOD
26	20130606	ONTARIO	SUDBURY, D	NOT APPLIC	NO	NOT APPLIC	YES	NOT APPLIC	26	60 TO 64 YE	FEMALE	MARRIED	NONE	NONE	NONE	SPOUSE / PA	2 PERSONS	EXCELLENT
27	20130606	ONTARIO	PEEL REGIO	NOT APPLIC	NO	IN PERSON	YES	NOT APPLIC	27	25 TO 29 YE	FEMALE	SINGLE/NEVE	NONE	NONE	NONE	OTHER	4 PERSONS	GOOD
28	20130606	ONTARIO	SUDBURY, D	NOT APPLIC	NO	NOT APPLIC	YES	NOT APPLIC	28	55 TO 59 YE	FEMALE	MARRIED	NONE	NONE	NONE	SPOUSE / PA	2 PERSONS	EXCELLENT
29	20130606	QUEBEC	RÉG. MONTE	NOT APPLIC	NO	NOT APPLIC	YES	NOT APPLIC	29	75 TO 79 YE	FEMALE	WIDOW/SEP	NONE	NONE	NONE	UNATTACHE	1 PERSON	GOOD
30	20130606	ONTARIO	PEEL REGIO	NOT APPLIC	NO	IN PERSON	YES	NOT APPLIC	30	25 TO 29 YE	FEMALE	SINGLE/NEVE	NONE	NONE	NONE	UNATTACHE	1 PERSON	VERY GOOD
31	20130606	BRITISH COI	FRASER SOU	FRASER	NO	NOT APPLIC	YES	NOT APPLIC	31	30 TO 34 YE	MALE	SINGLE/NEVE	NONE	NONE	NONE	UNATTACHE	1 PERSON	EXCELLENT
32	20130606	BRITISH COI	NORTH SHO	VANCOUVER	NO	NOT APPLIC	YES	NOT APPLIC	32	12 TO 14 YE	FEMALE	SINGLE/NEVE	NONE	NONE	NONE	CHILD 2 PAR	4 PERSONS	EXCELLENT
33	20130606	ONTARIO	WINDSOR-E	NOT APPLIC	NO	IN PERSON	NO	NO	33	65 TO 69 YE	MALE	MARRIED	NONE	NONE	NONE	SPOUSE / PA	2 PERSONS	EXCELLENT
34	20130606	NEW BRUNS	ZONE 3	NOT APPLIC	NO	NOT APPLIC	YES	NOT APPLIC	34	65 TO 69 YE	FEMALE	WIDOW/SEP	NONE	NONE	NONE	UNATTACHE	1 PERSON	VERY GOOD
35	20130606	YUKON/NWT	YUKON/NOR	NOT APPLIC	NO	ON TELEPHC	NO	NO	35	25 TO 29 YE	MALE	COMMON-LA	NONE	1 OR MORE	1 OR MORE	PAR SPOUSE	3 PERSONS	VERY GOOD
36	20130606	ALBERTA	CENTRAL ZC	NOT APPLIC	NO	NOT APPLIC	YES	NOT APPLIC	36	15 TO 17 YE	MALE	SINGLE/NEVE	NONE	NONE	NONE	CHILD 2 PAR	3 PERSONS	VERY GOOD
37	20130606	NOVA SCOTI	CAPE BRETC	NOT APPLIC	NO	IN PERSON	YES	NOT APPLIC	37	65 TO 69 YE	FEMALE	SINGLE/NEVE	NONE	NONE	NONE	UNATTACHE	1 PERSON	VERY GOOD
38	20130606	ONTARIO	HALTON REG	NOT APPLIC	NO	ON TELEPHC	YES	NOT APPLIC	38	12 TO 14 YE	MALE	SINGLE/NEVE	NONE	NONE	NONE	CHILD 2 PAR	3 PERSONS	VERY GOOD
39	20130606	ALBERTA	NORTH ZONE	NOT APPLIC	NO	IN PERSON	YES	NOT APPLIC	39	40 TO 44 YE	FEMALE	MARRIED	NONE	1 OR MORE	1 OR MORE	PAR SPOUSE	4 PERSONS	EXCELLENT
40	20130606	ONTARIO	NIAGARA RE	NOT APPLIC	NO	IN PERSON	YES	NOT APPLIC	40	50 TO 54 YE	MALE	MARRIED	1 OR MORE	NONE	1 OR MORE	PAR SPOUSE	3 PERSONS	GOOD
41	20130606	ONTARIO	WATERLOO	NOT APPLIC	NO	IN PERSON	YES	NOT APPLIC	41	25 TO 29 YE	FEMALE	MARRIED	NONE	NONE	NONE	SPOUSE / PA	2 PERSONS	EXCELLENT
42	20130606	MANITOBA	NORTHERN R	NOT APPLIC	NO	NOT APPLIC	YES	NOT APPLIC	42	15 TO 17 YE	FEMALE	SINGLE/NEVE	1 OR MORE	NONE	1 OR MORE	OTHER	5 OR + PER	EXCELLENT
43	20130606	ALBERTA	CALGARY ZC	NOT APPLIC	NO	NOT APPLIC	YES	NOT APPLIC	43	65 TO 69 YE	FEMALE	MARRIED	NONE	NONE	NONE	SPOUSE / PA	2 PERSONS	GOOD
44	20130606	ALBERTA	CALGARY ZC	NOT APPLIC	NO	IN PERSON	YES	NOT APPLIC	44	12 TO 14 YE	MALE	SINGLE/NEVE	NONE	1 OR MORE	1 OR MORE	CHILD 2 PAR	4 PERSONS	VERY GOOD
45	20130606	ALBERTA	CALGARY ZC	NOT APPLIC	NO	NOT APPLIC	YES	NOT APPLIC	45	40 TO 44 YE	FEMALE	MARRIED	NONE	1 OR MORE	1 OR MORE	PAR SPOUSE	4 PERSONS	VERY GOOD
46	20130606	ONTARIO	CITY OF TOR	NOT APPLIC	NO	IN PERSON	YES	NOT APPLIC	46	25 TO 29 YE	FEMALE	SINGLE/NEVE	NONE	NONE	NONE	OTHER	4 PERSONS	GOOD
47	20130606	ONTARIO	PEEL REGIO	NOT APPLIC	NO	IN PERSON	YES	NOT APPLIC	47	35 TO 39 YE	MALE	MARRIED	1 OR MORE	NONE	1 OR MORE	PAR SPOUSE	4 PERSONS	EXCELLENT

Nesstar Publisher v4.0.9 - cchs-2012.Nesstar (cchs_2012) - [My Projects]

File Edit Documentation Variables Data Publishing Tools Help

English

Projects:

- Dates of Collection
- Countries
- Geographic Coverage
- Geographic Unit
 - Unit of Analysis
 - Universe
 - Kind of Data
- Notes
- Time Method
- Data Collectors
- Frequency of Data Collection
- Sampling Procedure
 - Major Deviations from the Sample Design
 - Mode of Data Collection
 - Type of Research Instrument
 - Data Sources
- Characteristics of the Data Collection Situation
 - Actions to Minimize Losses
 - Control Operations
 - Weighting
 - Cleaning Operations
 - Notes
 - Response Rates
- Estimates of Sampling Error
- Other Forms of Data Appraisal
- Class or Status of the Study
- Locations of Data Collection
- Restrictions
- Access Authority
- Citation Requirement
- Deposit Requirement
- Access Conditions
- Disclaimer
- Notes
- Other Study Materials
- Datasets
 - cchs_2012
 - File Description
 - Key Variables & Relations
 - Variables
 - Data Entry
 - Cell Notes
 - Cube Setups
- Variable Groups
- Other Materials
- External Resources

Variables

Number	Name	Label	Width	StartCol	EndCol	Record	Decimals	Is Weight Variab
v1	VERDATE	Date of file creation	8	1	8	1		
v2	GEOGPRV	Province of residence of respondent-(G)	2	9	10	1	0	<input type="checkbox"/>
v3	GEOOPMF	Health Region - (G)	5	11	15	1	0	<input type="checkbox"/>
v4	GEODBCHA	B.C. Health Authority (BCHA) - (D)	4	16	19	1	0	<input type="checkbox"/>
v5	ADM_PRX	Health Component completed by proxy	1	20	20	1	0	<input type="checkbox"/>
v6	ADM_N09	Interview by telephone/in person	1	21	21	1	0	<input type="checkbox"/>
v7	ADM_N10	Respondent alone during interview	1	22	22	1	0	<input type="checkbox"/>
v8	ADM_N11	Answers aff./presence of another person	1	23	23	1	0	<input type="checkbox"/>
v9	ADM_RNO	Sequential record number	5	24	28	1	0	<input type="checkbox"/>
v10	DHHGAGE	Age - (G)	2	29	30	1	0	<input type="checkbox"/>
v11	DHH_SEX	Sex	1	31	31	1	0	<input type="checkbox"/>
v12	DHHGMS	Marital status - (G)	1	32	32	1	0	<input type="checkbox"/>
v13	DHHGLE5	No. pers. <= 5 years old in hhld - (D,G)	1	33	33	1	0	<input type="checkbox"/>
v14	DHHG611	No. persons 6-11 years in hhld - (D,G)	1	34	34	1	0	<input type="checkbox"/>
v15	DHHGL12	No. pers. < 12 years old in hhld - (D,G)	1	35	35	1	0	<input type="checkbox"/>
v16	DHHGLVG	Living arrangements/selected resp -(D,G)	2	36	37	1	0	<input type="checkbox"/>
v17	DHHGHSZ	Household size - (D, G)	1	38	38	1	0	<input type="checkbox"/>
v18	GEN_01	Self-perceived health	1	39	39	1	0	<input type="checkbox"/>
v19	GEN_02	Self-perceived hlth - compared 1 yr ago	1	40	40	1	0	<input type="checkbox"/>
v20	GEN_02A2	Satisfaction with life in general	2	41	42	1	0	<input type="checkbox"/>
v21	GEN_02B	Self-perceived mental health	1	43	43	1	0	<input type="checkbox"/>
v22	GEN_07	Perceived life stress	1	44	44	1	0	<input type="checkbox"/>
v23	GEN_08	Worked at job	1	45	45	1	0	<input type="checkbox"/>
v24	GEN_09	Self perceived work stress	1	46	46	1	0	<input type="checkbox"/>
v25	GEN_10	Sense of belonging - local community	1	47	47	1	0	<input type="checkbox"/>
v26	GENDHDI	Perceived Health - (D)	1	48	48	1	0	<input type="checkbox"/>
v27	GENDMHI	Perceived Mental Health - (D)	1	49	49	1	0	<input type="checkbox"/>
v28	GENGSWL	Satisfaction with life in general	1	50	50	1	0	<input type="checkbox"/>
v29	DOORG	Module: ORG - (F)	1	51	51	1	0	<input type="checkbox"/>
v30	ORG_1	Member of a voluntary organization	1	52	52	1	0	<input type="checkbox"/>
v31	ORG_2	Frequency - participate in activities	1	53	53	1	0	<input type="checkbox"/>
v32	DOSLP	Optional module: SLP - (F)	1	54	54	1	0	<input type="checkbox"/>

Documentation

Statistics

Include Weighted Statistics

Include Frequencies

List Missing At End

Sorting of Frequencies:

Summary Statistics Options:

Include Valid

Include Min

Include Max

Include Mean

Include Weighted Mean

Include StdDev

Include Weighted StdDev

Weights

Documentation

Value: Label:

Category Text:

Level Name:

GeoMap URI:

Variable information

Data Type:

Measure:

Is Time Variable

Getting Started with DDI (cont'd)

- Another tool: Colectica for Excel
 - Based in the US
 - Freeware
 - Documents variables and datasets directly from within Excel
 - Can be used to produce detailed (item-level) metadata for studies already completed
 - Creates metadata and documentation for surveys
 - Saves metadata directly in the Excel file
 - When the file is shared, so is the metadata

Getting Started with DDI (cont'd)

- Colectica for Excel
 - Allows direct importing and documenting of Stata data files (.dta) and SPSS data files (.sav)
 - Will automatically import variable names, labels and code lists from the data file
 - This metadata will be added to the documentation that is already stored
- Uses DDI Lifecycle (DDI version 3.1)
 - Allows us to manage the entire life cycle of a study as well as manage our legacy materials
 - Records audit trail through versioning

Getting Started with DDI (cont'd)

- Colectica Reader
 - Free tool
 - To view the metadata
 - No specialized software is needed
 - Generates documentation for variables and code lists
 - PDF, Word, HTML

Getting Started with DDI (cont'd)

- Other options available from Colectica
 - Colectica Portal
 - Allows you to browse, search and view the contents of the Repository on the web
 - Free
 - Colectica Repository
 - To store the metadata on a server
 - Not free

Colectica workspace

Tool bar options

The screenshot shows the Colectica Designer interface. At the top, there is a menu bar with 'Project', 'Import', 'Export', and 'Help'. Below the menu bar is a toolbar with various icons and dropdown menus. A red box labeled 'Tool bar options' points to the toolbar. The main workspace is divided into several sections. On the left, there is a 'Project' sidebar with a tree view showing 'Studies', 'Resource Packages', 'Metadata Citation', 'Conceptual', 'Classifications', 'Data Collection', and 'Data'. The 'Conceptual' section is expanded, showing 'Concepts', 'Universes', and 'Organizations'. The 'Classifications' section is also expanded, showing 'Categories', 'Codes', and 'Survey Instruments'. The 'Data' section is expanded, showing 'Data Collection Waves', 'Questions', 'Sequences', and 'Survey Instruments'. The main workspace displays a table of survey questions for the 'Los Angeles County Social Survey, 1997'. The table has columns for 'Name' and 'Question Text'. A red arrow points from the 'Tool bar options' box to the 'Add a Concept Set' option in the toolbar, and another red arrow points from the same box to the 'Add a Variable Set' option in the toolbar. A text box with the text 'This work space is used when entering info for DDI modules' is overlaid on the table.

Project Import Export Help

Save Languages Create a Checkout Add a Study Add a Series Add a Resource Package

Add a Concept Set Add a Universe Set Add an Organization Set

Add a Category Set Add a Code Set

Add a Data Collection Wave Add a Variable Set Add a Dataset Definition

Local Create Items Projects Conceptual Classifications Data Collection Data Modules

Project

Studies

Resource Packages

Metadata Citation

Conceptual

Concepts

Universes

Organizations

Classifications

Categories

Codes

Data Collection

Data Collection Waves

Questions

Sequences

Survey Instruments

Data

Project

Repository Search

Checkouts

Los Angeles County Social Survey, 1997 Expand Agency us.datapass.uclassda Revision 2

Last Updated 2012-02-18 01:06

View DDI Create Documentation Revision History

Existing questions...

Name	Question Text
R10C	Random number for ordering Q
CON2	I would like to interview someone in your household. In order to determine who to interview, I need to know how many people
CON3	Of the people you just included in your count of household members, what is the first name of the one who will have the next
CON4	What is the first name of the person in your household who had the last birthday?
CON5	In that case, it will take me a moment to select a person to interview based on a random selection table I have. What is the fi
CLIK	Before we start, I would like to assure you that this interview is confidential and completely voluntary. If we should come to a
SEX	Are you male or female?
LA_LAST5	I have a couple questions about how things are in the Los Angeles area. In the last five years do you think that Los Angeles h
ETHGRPLA	Has the wide variety of ethnic groups in Los Angeles helped or hurt the quality of life here?
CONFLICT	Now, I'll ask you some questions about ethnic relations in Los Angeles. Are ethnic groups in conflict in Los Angeles/Southern C

CSID Expand Agency us.datapass.uclassda Revision 1

Last Updated 2012-02-16 11:08 View DDI Create Documentation Revision History

Question Instructions External Aids Conceptual Extended

Question Text

Case ID

This work space is used when entering info for DDI modules

Example codebook from Colectica (PDF)

Bookmarks

- Cover
- Higher Education General Information Survey (HEGIS), 1968-1969: Financial Statistics [Computer file].
 - Overview
 - Abstract**
 - Purpose
 - Sample
 - Study Citation
 - fin_merge_1969
 - Citation
 - Record: fin_merge_1969
 - FICE - FICE CODE
 - OESTATE - OE STATE CODE
 - SEQUENCE - SEQUENCE NUMBER
 - SYSCODE - SYSTEM CODE
 - SEQCODE - SEQUENCE CODE
 - STATCOD - STATUS CODE
 - OECODE - OLD CODE PRIOR ASSIGN FICE
 - GEOGCODE - GEOGRAPHIC CODE

Abstract

The Higher Education General Information Survey (HEGIS) series was designed to provide comprehensive information on various aspects of postsecondary education in the United States and its territories (American Samoa, Guam, Puerto Rico, the Virgin Islands, and the Marshall Islands) and Department of Defense schools outside the United States. Data are available for both public and private two-year and four-year institutions. The HEGIS Financial Statistics component for 1968-1969 offers financial data for 2,814 institutions of higher education. Variables include current funds revenues by source, current funds expenditures by function, physical plant assets, indebtedness on physical plant, endowment assets, and changes in fund balances.

Purpose

Sample

Study Citation

Alternate Title: HEGIS

Creator: United States Department of Education. National Center for Education Statistics.

Publisher: Inter-university Consortium for Political and Social Research [distributor].

Copyright: doi:10.3886/ICPSR02094.v1



Getting Started with DDI (cont'd)

- Remember the two types of DDI?
 - DDI Codebook (DDIC) and DDI Lifecycle (DDIL)
 - Two separate programs
- Challenge
 - How to go from DDIC to DDIL?
 - Even though DDIL is fast becoming the more preferred one, if you are already using DDIC, it would be too much work to convert to DDIL
 - Insufficient resources to do this conversion for most institutions (people, \$)

Getting Started with DDI (cont'd)

- The latest news
 - Nesstar has created a new API
 - Hopefully be able to publish DDI Lifecycle to DDI Codebook on the Nesstar server
 - This integration has been implemented between Nesstar and Colectica
 - Will allow users greater opportunities for
 - Expanded metadata production
 - Improved data publication and visualization



Next ...

- We have
 - learned the history of DDI
 - seen who is using it
 - learned how it can streamline our research
 - seen some tools that will help us use it
- Next, we will look at examples of DDI in 2 tools
 - These are to show you how important it is to have tools that will interpret it for you so you don't have to know XML

Examples of using DDI

- Check out how Nesstar Webview works
 - Video Tutorial 1: Accessing Health of Houston Survey Data (2 min)
 - <http://www.youtube.com/watch?v=NNvcS1xsOEg>
 - Video Tutorial 2: Looking at Survey Variables in Nesstar (3 min)
 - http://www.youtube.com/watch?v=xkBGC-5K_Bk
 - Video Tutorial 3: Using the Tabulation Function in Nesstar (4 min)
 - <http://www.youtube.com/watch?v=qqTlN5mVRdg>

Examples of using DDI (cont'd)

- Check out how Colectica works
 - 11 videos to choose from
 - all are 2 minutes or less
 - <http://www.youtube.com/user/Colectica>
 - How to Open an Existing Study
 - User Interface Overview
 - Export DDI 3
 - Generate Documentation

Back in 15!





Review

- We have learned
 - the history of DDI
 - who is using it
 - how it can streamline our research
- We have seen
 - 2 tools and examples
 - Nesstar and Colectica for Excel
- Remember
 - There are 2 different lines of DDI development
 - DDIC and DDIL



Data Lifecycle Workflow

- Integrating Metadata and Research Data Management into a data lifecycle workflow
 - What metadata
 - When or where to integrate it
 - What does research data management have to do with this



Data Lifecycle Workflow (cont'd)

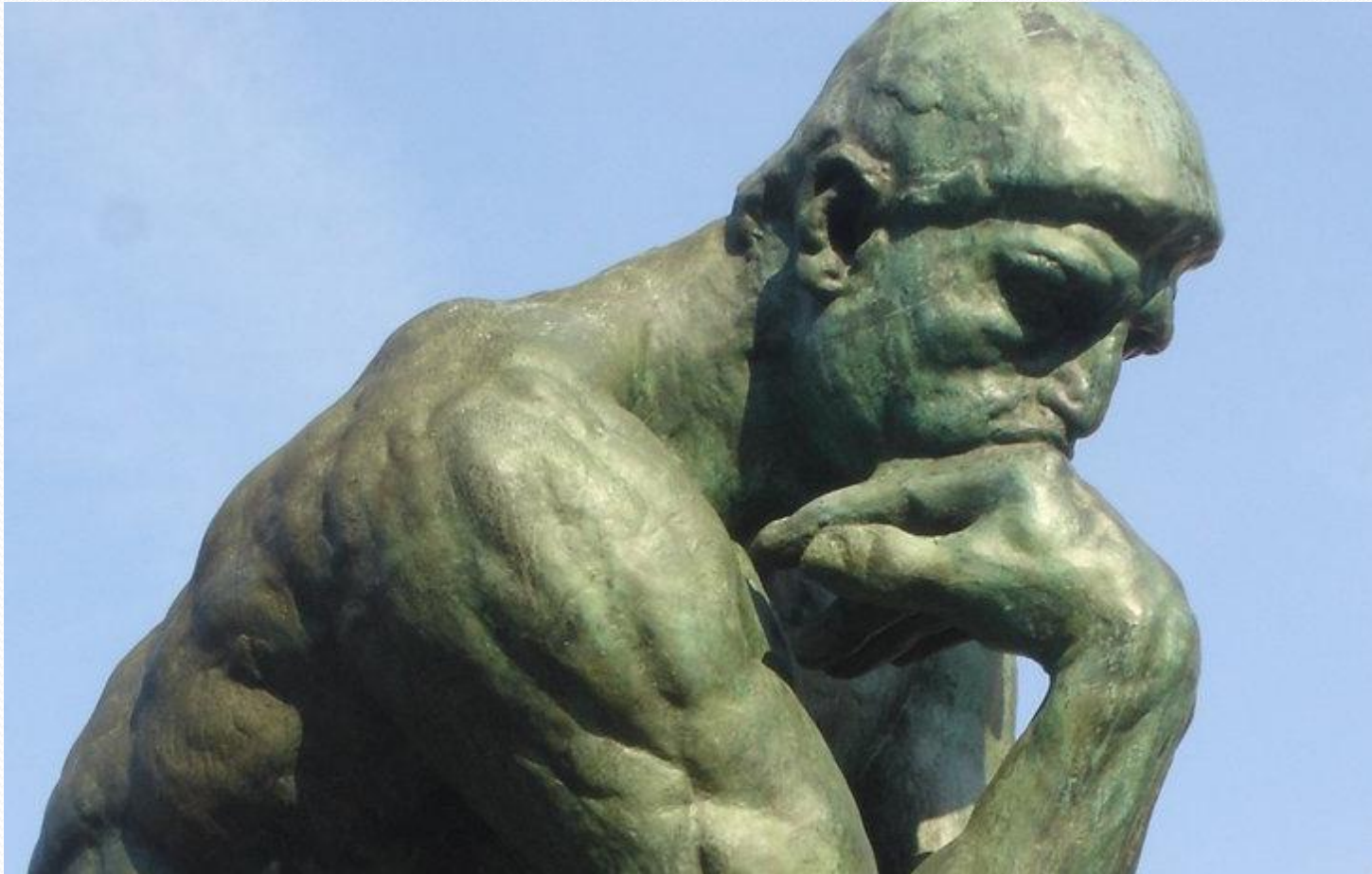
- DDI Metadata accompanies and enables
 - Data conceptualization
 - Collection
 - Processing
 - Distribution
 - Discovery
 - Analysis
 - Repurposing
 - Archiving



Data Lifecycle Workflow (cont'd)

- Remember
 - More efficient and easier to capture information about all stages of the research workflow at the time of its occurrence rather than after the fact

Time to think!

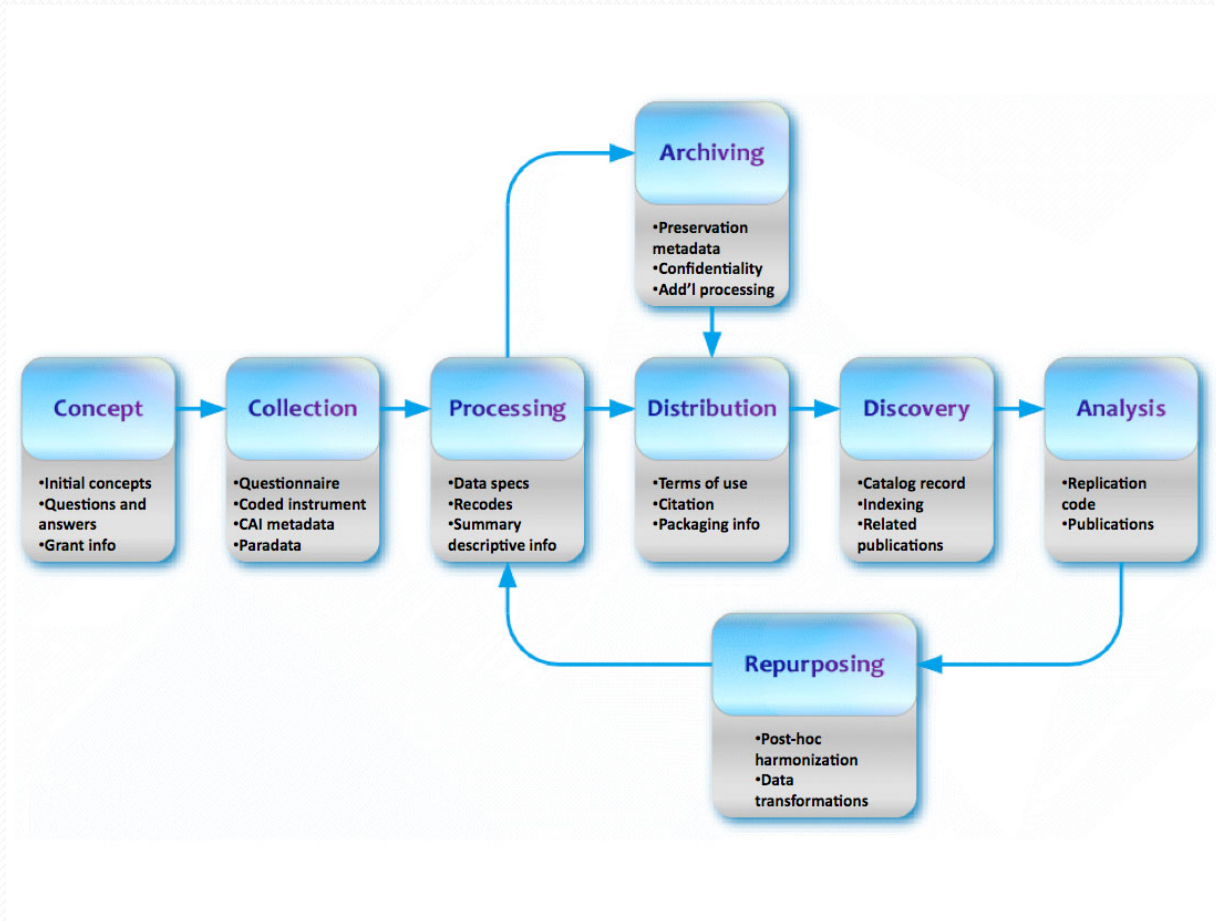




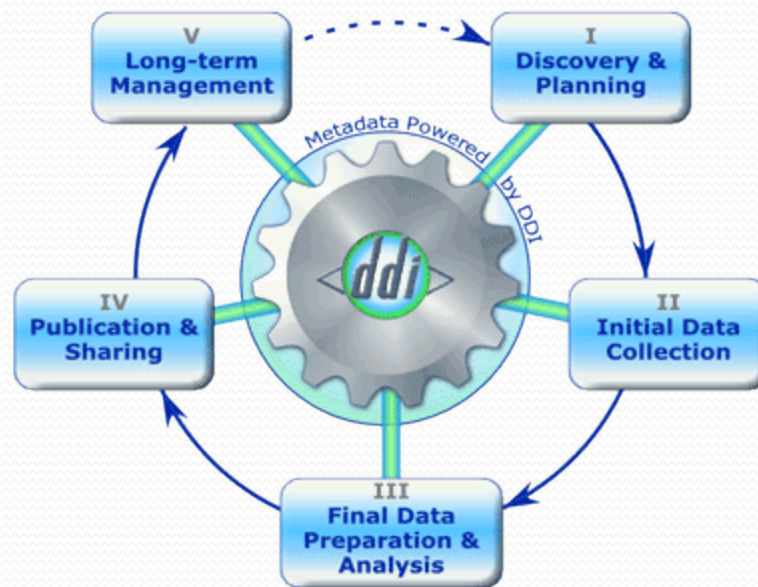
Exercise 2

- *When or where* would you like to see metadata integrated into a data lifecycle workflow?
- What does research data management have to do with this?

DDI Research Lifecycle



DDI Lifecycle 3.2



<http://www.ddi-alliance.org/>





Workshop Materials

- To access materials from this workshop
 - They will be on the NADDI Conference website
 - They are also here:

<http://www.library.carleton.ca/content/na-ddi-north-american-ddi-user-conference>



References

- DDI Alliance. *<ddi> Data Documentation Initiative*. Retrieved March 17, 2014 from <http://www.ddi-alliance.org/>
- GNU. *GNU operating system*. Retrieved March 18, 2014 from <http://www.gnu.org/>
- Iverson, J. & Stephenson, E. (2013). *DDI-Lifecycle and Colectica at the UCLA Social Science Data Archive*. Presentation at the North American Data Documentation Conference (NADDI) 2013. <http://hdl.handle.net/1808/11049>
- Vardigan, M. & Wackerow, J. (2013). DDI – A metadata standard for the community. Paper presented at the North American Data Documentation Initiative Conference (NADDI) 2013. <http://hdl.handle.net/1808/11056>
- W3C. *Extensible Markup Language (XML)*. Retrieved March 17, 2014 from <http://www.w3.org/XML/>



Contact Information

Jane Fry

Data Specialist

MacOdrum Library

Carleton University

jane.fry@carleton.ca

[janefry@gmail.com](mailto:janfry@gmail.com)

613.520.2600 x1121