

Bringing it together

Jon Johnson

12 September 2019

CLOSER, UCL Institute of Education



Overview

- Bringing it all together
- Data management problem
- Infrastructures
- Software and Tooling
- Controlled Vocabularies & XKOS



What does your data look like?

- Codebook can describe
 - Unit and dimensional data
 - Basic capture information (questions, derivation codes, secondary use of source data)
 - Archival data file structures – limited relational information
 - Focus is on the individual study/data set
- Lifecycle can describe
 - Questionnaire structure
 - Multi-wave studies and their internal relationships
 - Cross study relationships
 - Common conceptual material



Infrastructure profiles: DDI Codebook

- Identification
 - Only requires instance identifier
 - Content is nested limiting the need for internal references
 - Uses standard ID and IDRef (supported by standard XML validation tools)
 - Supports capture of standard DDI URN
- XML only binding
- Current tools for creation of an instance, catalog of instances, and transformation to PDF document or web site
- Assumes XML instance is a publication in itself and will be managed as such



Infrastructure profiles: DDI Lifecycle

Identification

- Required by most classes
- Registry of DDI Agent identification
- Uses DDI structure for identification that resolves to a URN (requires secondary validation tool)
- Allows and encourages reuse of metadata between instances

XML only binding

Current tools for creation of an instance, repository of DDI objects, transformation to PDF document or web site, and creation of questionnaires

Active management of metadata



DDI Tools

Tools	Version supported	
	DDI Codebook	DDI Lifecycle
Colectica	Yes	Yes
Nesstar	Yes	No
Archivist	No	Yes
Questionnaire Design * Development Tool	No	Yes
NADA (World Bank)	Yes	No
World Bank Publisher*	Yes	?
Dataverse.org	Yes	?
Questionnaire crosswalks (redCap, LimeSurvey, Blaise)	No	Yes
Statistical packages (StatTransfer, Sledgehammer)	Yes	Yes



Controlled Vocabularies coverage: 2010 -

- Controlled Vocabularies (CVs) are created by the DDI to support commonly used vocabularies among DDI users and CESSDA members
- CVs are published at
 - <http://www.ddialliance.org/controlled-vocabularies>
- Currently published in:
 - Customized Genericcode format (XML)
 - XLS spreadsheet
 - HTML version for viewing
- CVs may be used by any version of DDI or other standard



Extending the standard

Attribute Pairs

- Sometimes you just need something that the standard does not support and you probably don't want to share it, but you want to just tweak something for your own use, or while the standard catches up.
- UserAttributePair is a Key:Value Pair which can add an element or refine an existing element
 - The Quality Statement in DDI could be refined by a controlled vocabulary
 - In DDI3.2 the name of a question (CAI name) was not available, so we used a bespoke AttributePair to hold that – we can now move that into the new DDI3.3 QuestionName element



Capturing Computations

- In addition to In/Out parameters and Binding
 - Human readable description
 - Language: Pseudocode, SPSS, Stata, Python etc
 - Command

Gives you the capability to

- Specify a computation
- Execute a specific piece of code
- Generate new code in many language parsed from the pseudocode

Can be used for derivations, inline actions (during data collection), processing pipelines etc



DDI RDF Vocabularies


- XKOS* standardizes the representation of statistical classifications as linked metadata
 - It defines a number of terms that enable the representation of statistical classifications with their structure and textual properties, as well as the relations between classifications.
 - Refines SKOS semantic properties to allow the use of more specific relations between concepts, builds upon the SKOS W3C Recommendation and implements the Neuchâtel and GSIM statistical models
- Disco
 - Supports the discovery of microdata sets and related metadata using RDF technologies in the Web of Linked Data

* <http://rdf-vocabulary.ddialliance.org/xkos.html>



NADA – Codebook publishing


International Housing Survey Network



International Household Survey Network

HOMEMICRODATA CATALOGCITATIONS

HOME › CENTRAL DATA CATALOG › SLV_1992_PHC_V01_M_V02_A_IPUMS



5th Census of Population 1992 - IPUMS Subset
El Salvador, 1992
General Directorate of Statistics and Censuses, Minnesota Population Center
Last modified March 23, 2019 • Page views 13140 • [Study website](#) • [Metadata](#) [CODEBOOK](#) [JSON](#)

STUDY DESCRIPTIONDOCUMENTATIONDATA DESCRIPTIONGET MICRODATARELATED PUBLICATIONS

Identification

Version

Scope

Coverage

Producers and sponsors

Sampling

Data Collection

Data access

Disclaimer and copyrights

Contacts

Metadata production

Identification

IDHO
SLV_1992_PHC_v01_M_v02_A_IPUMS

TITLE
5th Census of Population 1992 - IPUMS Subset


COUNTRY	
Name	Country code
El Salvador	SLV

ABSTRACT
IPUMS-International is an effort to inventory, preserve, harmonize, and disseminate census microdata from around the world. The project has collected the world's largest archive of publicly available census samples. The data are coded and documented consistently across countries and over time to facilitate comparative research. IPUMS-International makes these data available to qualified researchers free of charge through a web dissemination system.

The IPUMS project is a collaboration of the Minnesota Population Center, National Statistical Offices, and international data archives. Major funding is provided by the U.S. National Science Foundation and the Demographic and Behavioral Sciences Branch of the National Institute of Child Health and Human Development. Additional support is provided by the University of Minnesota Office of the Vice President for Research, the Minnesota Population Center, and Sun Microsystems.

KIND OF DATA
Census/enumeration data [oen]

CLOSER (LMIC) Atlas



The home of longitudinal research

Microdata Library

Date range1934 – 2021

Data access

- Data available
- Data not available



Lifecycle publishing – multi-study

closer
discovery



8
Studies



246
Data Files



79,412
Variables



214
Questionnaires



29,231
Questions



Questions, variables & lineage

1970 British Cohort Study

Age 42 Survey (2012)

BCS70 Paper Self Completion (2012) Dataset

Variable

Details

Lineage

BCS70 Paper Self Completion (2012) Dataset - BD9AUDG

(Derived) AUDIT-PC Group

Derived Variable

BCS70 Paper Self Completion (2012) Dataset - BD9AUDIT

(Derived) Total AUDIT-PC Score

Total Score

BCS70 Paper Self Completion (2012) Dataset - B9SCQ32

PAPI:Frequency of having an alcoholic drink

Item Score 1

qi_32

How often do you have a drink containing alcohol?

Question 1

BCS70 Paper Self Completion (2012) Dataset - B9SCQ33

PAPI:Number of alcoholic drinks consumed on a typical day when drinking

Item Score 2

qi_33

How many drinks containing alcohol do you drink on a typical day when you are drinking?

Question 2

BCS70 Paper Self Completion (2012) Dataset - B9SCQ34

PAPI:Frequency of not being able to stop drinking once started over last year

Item Score 3



DDI Lifecycle in Archives



Find Social and Economic Research Data

1320 results found

Filter summary

Advanced search

Reset filters

Clear search

▼ Topic



Search topics

Results per page

30

Sort by

Relevance



1

2

3

4

...



► Collection years



► Country



► Publisher



▼ Language of data files



x fi (1320)

[Finnish Youth Survey 2003](#)

Advisory Council for Youth Affairs (Nuora); Finnish Youth Research Society. Finnish Youth Research Network

The main theme of the survey was entrepreneurship. First, the respondents were presented a set of attitudinal statements about work, working life, unemployment and success at work. They were asked what kind of company they would set up, if it was economically possible. The survey also charted whether there were entrepreneurs in the family of the respondent. The respondents were asked to assess how closely they identified with different groups or areas, such as family, school or work community...

[Living Conditions of Social Assistance Recipients and Social Work 1987](#)

*Ritakallio, Veli-Matti (University of Turku. Department of Social Policy);
Salavuo, Kari (University of Turku. Department of Social Policy)*

The survey charted living conditions, experiences of life and client relationship of social assistance recipients and social work. Firstly, respondents were asked about previous contacts with social services office, previous occupation and job and length of and reason for unemployment. They were also asked to state different illnesses and ailments they had suffered from in the past twelve months. The survey also studied respondents'



Lifecycle publishing – single study

MIDUS

Search

Explore

Basket 0

Survey

Administration

Recession Experience

Health

Education, Occupation, and Marital Status

Household Roster and Children

Caregiving

Living Arrangements

Race and Ethnicity

Life Satisfaction

Your Health

Self-Evaluated Health

Life Satisfaction

Change in Health Status

Change in Health Expectations

Comparative Health/Memory Items

Locus of Control

Somatic Amplification

Physical Symptoms Past 30 Days

Chronic Medical Conditions

Prescription Medicine Used

Non-Prescription Medicine

Dietary Supplements

	M1P1	M2P1	M3P1	MKE1
<div></div> Life Satisfaction (5-items)	A1SSATIS	B1SSATIS	C1SSATIS	BACSATIS
<div></div> Life Satisfaction (6-item version)		B1SSATIS2	C1SSATIS2	BACSATIS2



Lifecycle publishing – single study

Concept  Life Satisfaction (6-item version)

Comparability Class Constructed Item - Difference exists among waves because some items were constructed only in particular waves.

Comparability Notes Not constructed in M1.

tatistics

[Code Comparison](#)

[Correspondence Tree](#)

Life Satisfaction (6-item version)



Question & Answer Session

Jon Johnson

12 September 2019

CLOSER, UCL Institute of Education



Thank you

Contact details



@CLOSER_UK

closer@ucl.ac.uk

j.Johnson@ucl.ac.uk

h.mills@ucl.ac.uk

Resources

closer.ac.uk

discovery.closer@ucl.ac.uk

closer.ac.uk/wiki

ddialliance.org