



INSTITUTE FOR SOCIAL RESEARCH • SURVEY RESEARCH CENTER
SURVEY RESEARCH OPERATIONS
UNIVERSITY OF MICHIGAN

Metadata from Blaise and DDI 3.0/3.2

Gina Cheung
Beth-Ellen Pennell

North American DDI Conference
April 1-2, 2014



Agenda

- Blaise Metadata (MQDS)
- DDI 3.0/3.1 – DDI 3.2
- Next Step



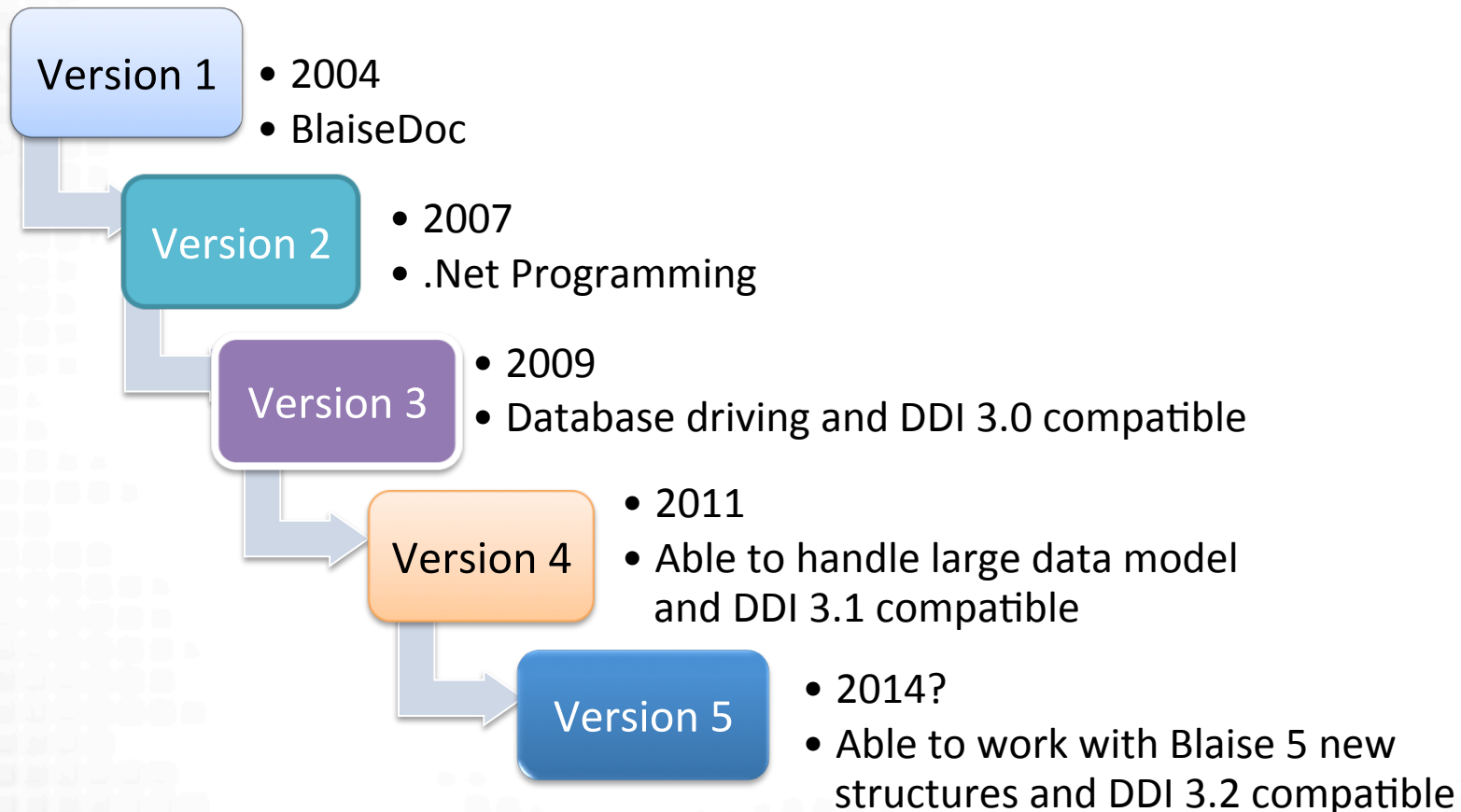
What is MQDS?

- The Michigan Questionnaire Documentation System (MQDS) was designed to extract comprehensive metadata from **Blaise** survey instruments and render it as an eXtended Markup Language (XML) document using the Data Documentation Initiative (DDI) standard.

From 2013 NADDI presentation



MQDS Development Timeline



From 2013 NADDI presentation



MQDS usage

- **Free** to all Blaise community users
- Questionnaire design
- Questionnaire documentation
- Quick data transformation (BDB format to SAS, SPASS or SQL DB format)
- Creating interviewers' training materials
- Codebook with universe info and “go-to” logic



MQDS Limitations

- Complex Blaise questionnaire roster structure
- Blaise alien router
- Paradata documentation
- Not a tool for Survey life circle documentation yet....

From 2013 NADDI presentation



DDI History

- **DD2:**
 - Documents variables in a dataset
 - Question text from a survey is associated with a single variable
 - No detailed description of the question or response options
 - No question flow, only simple forward and backward links
- **DDI 3.0/3.1**
 - Describes instrument flow
 - Textual description of conditions for if/then, loops, and dynamic text

From 2013 EDDI Colectica's presentation



Blaise Metadata Show Case

- Collaborative Psychiatric Epidemiology Surveys (CPES): to collect data re: distributions, correlates and risk factors of mental disorders as well as explore ethnic disparities, support systems, discrimination, and assimilation
 - Interviewing extended from 2001 -- 2003
 - All probability samples of non-institutionalized adults living in coterminous US
 - Version of World Health Organization's Composite International Diagnostic Interview (CIDI) programmed in Blaise
- CPES Harmonization Project:
 - Harmonize and merge CPES datasets
 - Prepare enhanced tools (interactive codebook) and data for on-line analysis
 - Develop Web site
 - Disseminate for public use

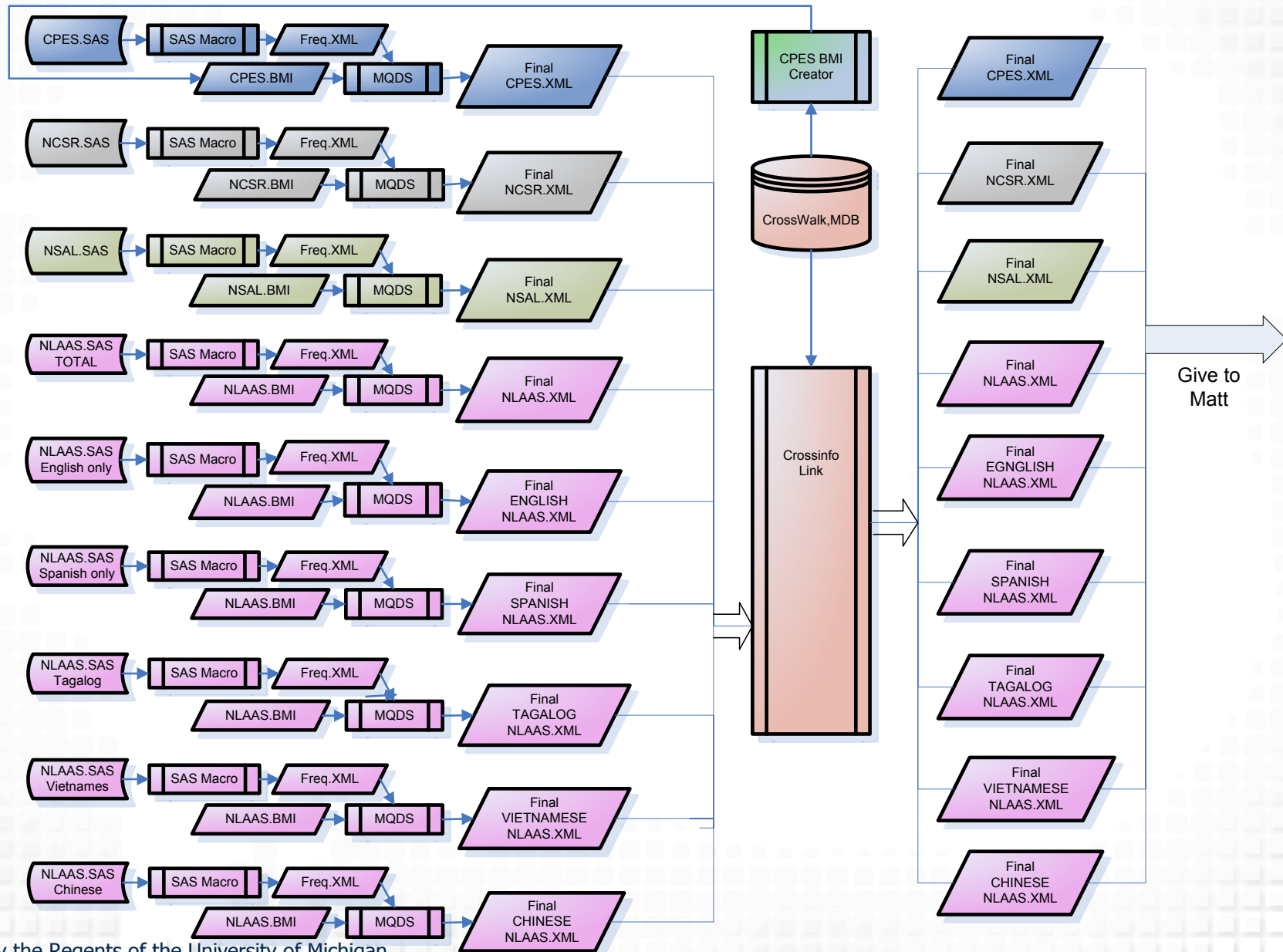


Challenges

- Documented 3 instruments with multiple datamodels
 - Needed to create pseudo instrument for CPES merged dataset
 - Edited instrument datamodels to improve documentation
- Created eXtended Markup Language (XML) files with question metadata for 9 XML files
 - CPES
 - Individual studies
 - NLAAS five languages




Complete Process





Interactive Codebook

Collaborative Psychiatric Epidemiology Surveys



Home
About CPES
Training Resources
Using CPES
Interactive Codebook
Online Analysis
FAQs

Announcements

- Restricted Data Now Available for NSAL, 2001-2003
- Additional Videos of 2008 Workshop
- New Restricted Files Now Available
- Video Tutorials from 2008 Workshop Released
- Assistance with the CPES and NSAL Datasets

Welcome to CPES

The National Institute of Mental Health Collaborative Psychiatric Epidemiology Surveys (CPES) provides data on the distributions, correlates, and risk factors of mental disorders among the general population, with special emphasis on minority groups. This project joins together three nationally representative surveys: the [National Comorbidity Survey Replication \(NCS-R\)](#), the [National Longitudinal Survey of Youth \(NLSY\)](#), and the [National Latino and Asian American Study \(NLAAS\)](#).

See the [Web Site User's Guide](#) in Using CPES.

Compare Questions & Variables

CPES is a harmonization of three surveys. You can view the codebook, by subject, or individual survey.

Browse Interactive Codebook

Browse by Subject

| DSM-IV | ICD-10 |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <ul style="list-style-type: none">• Mood• Anxiety• Substance• Impulse Control | <ul style="list-style-type: none">• Mood• Anxiety• Substance• Impulse Control |

[Home](#) > [Interactive Codebook](#) > [CPES](#)

Interactive Codebook

You can also choose to browse by selecting [DSM-IV/ICD subjects](#).

CPES | NCSR | NLAAS | NSAL

Below is a list of all the sections in CPES. They are currently sorted alphabetically by section name. You can also choose to [view them in the order they appeared](#) in the questionnaire.

The [PDF codebook](#) can answer many basic questions about the data; you will need a MyData account to download the codebook.

- [30-Day Functioning](#)
- [30-Day Symptoms](#)
- [Acculturative Stress](#)
- [Adult Demographics](#)
- [Agoraphobia](#)
- [Attention Deficit Hyperactivity Disorder](#)
- [Childhood](#)
- [Childhood Demographics](#)
- [Children](#)
- [Chronic Conditions](#)
- [Conduct Disorder](#)
- [Constructed Demographic Variables](#)
- [Context of Exit Questions](#)
- [Demographics](#)
- [Depression](#)
- [Discrimination](#)
- [DX Adult Separation Anxiety Disorder](#)
- [DX Agoraphobia](#)
- [DX Alcohol Abuse](#)
- [DX Alcohol Dependence](#)
- [DX Anorexia Nervosa](#)
- [DX Attention Deficit Disorder](#)
- [DX Binge Eating Disorder](#)
- [DX Bipolar](#)



CPES V04454 Frequencies

Home > Interactive Codebook > CPES > 30-Day Functioning

30-Day Functioning

The table below only lists variables in the public use file. If you want to see the position and frequency distribution of variables that are in the **restricted** file, you can [view all variables in this section](#).

* indicates there is some missing data
 † This variable appears in different sections from CPES.

| ID | Label |
|--------|----------------------------------------------------------------------------------------|
| V04445 | Status of overall functioning |
| V04446 | Amount of change in overall functioning |
| V04447 | Frequency of experiencing physical symptoms |
| V04448 | Severity of physical symptoms |
| V04449 | Overall mental health |
| V04451 | Amount of change in overall mental health |
| V04452 | # days in past 30 days with no activities due to emotions/nerves/ment health/subst use |
| V04454 | No activities due to emotions/nerves/ment health/subst use |
| V04455 | # days no activities due to emotions/nerves/ment health/subst use |
| V04453 | INTERVIEWER C |
| V04456 | # days past mth |
| V04457 | INTERVIEWER C |

Home > Interactive Codebook > CPES > 30-Day Functioning > V04454

<-- previous variable next variable -->

CPES V04454 NCSR FD4A NLAAS FD4A NSAL C29C †

† For the individual studies, this variable exists in different sections from CPES.

Variable Label: No activities due to emotions/nerves/ment health/subst use

FD4a

Was that due to your emotions, nerves, mental health, or your use of alcohol or drugs?

1 YES
 5 NO

| Value | Label | Frequency | Valid Percent |
|-------|---------|-----------|---------------|
| 1 | YES | 111 | 2.14% |
| 5 | NO | 356 | 7.00% |
| . | Missing | 19546 | 39.14% |

Disclaimer: Frequencies displayed above are not weighted.

- Missing Data Codes: . (Missing)
- Total Cases: 20013

You can also [view this variable in SDA](#).

Home > Interactive Codebook > NSAL > Psychological Resources and Health > C29C

<-- previous variable next variable -->

CPES V04454 NCSR FD4A NLAAS FD4A NSAL C29C †

† For the individual studies, this variable exists in different sections from CPES.

Variable Label: No activities due to emotions/nerves/ment health/subst use

C29c

Was that due to your emotions, nerves, mental health, or your use of alcohol or drugs?

1 YES
 5 NO

[View Universe](#)

| Value | Label | Frequency | Valid Percent | Total Percent |
|-------|---------|-----------|---------------|---------------|
| 1 | YES | 21 | 14.69% | 00.35% |
| 5 | NO | 122 | 85.31% | 02.01% |
| . | Missing | 5939 | - | 97.65% |

Disclaimer: Frequencies displayed above are not weighted.

- Missing Data Codes: . (Missing)
- Total Cases: 6082

You can also [view this variable in SDA](#).



ICPSR Developed Web site

- Used XML files and the crosswalk to create interactive codebook
- Provided
 - keyword search capability
 - downloadable datasets and codebooks
 - online analysis tools
 - links to study documentation, publications, and related sites
- Web site maintenance and user support

| Table 3.3: ICPSR Reports of Usage of SRC Data Sets Archived -- Downloads | | Downloads | Downloads | Publications |
|------------------------------------------------------------------------------------------------------------------------|--------------|--------------|----------------|--------------|
| Study Name | Study No. | 1996-2007 | 2008-2013 | |
| Individual Studies | | | | |
| Alameda County [California] Health and Ways of Living Study, 1999 Panel | 4432 | 491 | 3,465 | 8 |
| American Panel Study: 1956, 1958, 1960 | 7252 | 620 | 598 | 80 |
| Americans' Changing Lives: Waves I, II, III, and IV, 1986, 1989, 1994, and 2002 | 4690 | 800 | 17,430 | 274 |
| Census of Population and Housing, 1990 [United States]: Extract Data | 2889 | 2,679 | 2,056 | 2 |
| China Study, 1964 | 7251 | 205 | 445 | 3 |
| Chitwan Valley [Nepal] Family Study: Changing Social Contexts and Family Formation | 4538 | 937 | 21,736 | 69 |
| Collaborative Psychiatric Epidemiology Surveys (CPES), 2001-2003 [United States] | 20240 | 2,340 | 135,063 | 459 |
| Communities in Charge Survey, 2001-2003 [Alameda County, California, Austin, Texas, and Southern Maine] | 4638 | 16 | 115 | 5 |
| County Chairmen Study, 1965 | 7200 | 33 | 84 | 0 |
| Detroit Arab American Study (DAAS), 2003 | 4413 | 663 | 5,500 | 2 |
| Dynamics of Economic and Demographic Behavior: Clean Processes From the Panel Study of Income Dynamics (PSID) (S01239) | 1239 | 54 | 1,183 | 5 |
| Financial Affairs Study, October 1953 | 7220 | 30 | 104 | 0 |
| Foreign Affairs Study, June 1951 | 7219 | 103 | 128 | 1 |
| Intergenerational Study of Parents and Children, 1962-1993: [Detroit] (S09902) | 9902 | 8,887 | 7,158 | 36 |
| Jobs II Preventive Intervention for Unemployed Job Seekers, 1991-1993: [Southeast Michigan] (S02739) | 2739 | 674 | 6,591 | 1 |
| Multi-City Study of Urban Inequality, 1992-1994: [Atlanta, Boston, Detroit, and Los Angeles] (S02535) | 2535 | 9,766 | 12,853 | 77 |
| National Comorbidity Survey: Baseline (NCS-1), 1990-1992 (S06693) | 6693 | 28,674 | 17,346 | 377 |
| National Survey of Black Americans, Waves 1-4, 1979-1980, 1987-1988, 1988-1989, 1992 (S06668) | 6668 | 15,502 | 8,968 | 79 |



New to DDI 3.2

- The main new things around questions and surveys in DDI 3.2 are:
 - Fragments, Groups, Represented variables
 - Modeling data flows through an Instrument with InParameters, OutParameters, and Bindings
 - QuestionGrids
 - QuestionBlocks
 - Scale questions
 - Ranked questions
 - Point distribution questions



Fragments

DDI 3.0

```
<s:StudyUnit id="GITFIELD"> ...  
<r:Citation> ...  
<s:Abstract id="ABS_1"> ...  
<r:UniverseReference isReference="true"> ...  
<s:Purpose id="PUR_1"> ...  
<c:ConceptualComponent isMaintainable="true"  
  id="CC_1"> ...  
<d:DataCollection isMaintainable="true" id="DC1"  
  version="1.0.0"  
  versionDate="2014-03-26"> ...  
<l:LogicalProduct id="LP1"> ...  
  <l:Category isVersionable="true"  
    id="cat_17_C82" missing="false">  
    <l:CategoryName>Male</  
l:CategoryName>  
    <r:Label xml:lang="en-US">Male</  
r:Label>  
  </l:Category>  
  ...
```

DDI 3.2

```
<ddi:Fragment>  
  <l:Category>  
    <r:Agency>int.example</r:Agency>  
    <r:ID>Male</r:ID>  
    <r:Version>1</r:Version>  
    <r:Label>  
      <r:Content xml:lang="en-US">Male</  
r:Content>  
    </r:Label>  
  </l:Category>  
</ddi:Fragment>
```




Groups

DDI 3.1

```
<l:VariableGroup id="varGroup1">
  <r:UserID type="G">GITFIELD</r:UserID>
  <l:GroupType>section</l:GroupType>
  <r:Label xml:lang="English">GITFIELD</r:Label>
  <l:VariableReference>
    <r:ID>V1</r:ID>
  </l:VariableReference>
  <l:VariableReference>
    <r:ID>V461</r:ID>
  </l:VariableReference>
  <l:VariableReference>
    <r:ID>V23</r:ID>
  </l:VariableReference>
</l:VariableGroup>
```

DDI 3.2

```
<l:VariableGroup>
  <r:Agency>UoM</r:Agency><r:ID>VG1</r:ID>
  <r:Version>1</r:Version>
  <l:VariableGroupName>
    <r:String>varGroup1</r:String>
  </l:VariableGroupName>
  <r:VariableReference>
    <r:Agency>UoM</r:Agency><r:ID>V1</r:ID>
    <r:Version>1</r:Version>
    <r:TypeOfObject>Variable</r:TypeOfObject>
  </r:VariableReference>
  <r:VariableReference>
    <r:Agency>UoM</r:Agency><r:ID>V461</r:ID>
    <r:Version>1</r:Version>
    <r:TypeOfObject>Variable</r:TypeOfObject>
  </r:VariableReference>
  <r:VariableReference>
    <r:Agency>UoM</r:Agency><r:ID>V23</r:ID>
    <r:Version>1</r:Version>
    <r:TypeOfObject>Variable</r:TypeOfObject>
  </r:VariableReference>
</l:VariableGroup>
```



Represented Variables

```
<ddi:Fragment>
  <l:Variable>
    <r:Agency>int.example</r:Agency>
    <r:ID>lvar1</r:ID>
    <r:Version>1</r:Version>
    <l:VariableName>
      <r:String>marStat2000</r:String>
    </l:VariableName>
    <r:Label>
      <r:Content>Marital status of respondent in 2000</r:Content>
    </r:Label>
    <l:RepresentedVariableReference>
      <r:Agency>int.example</r:Agency>
      <r:ID>repVar1</r:ID>
      <r:Version>1</r:Version>
      <r:TypeOfObject>RepresentedVariable</r:TypeOfObject>
    </l:RepresentedVariableReference>
    <l:VariableRepresentation>
      <r:CodeRepresentation/>
    </l:VariableRepresentation>
  </l:Variable>
</ddi:Fragment>
```



Data flow: InParameters, OutParameters, and Bindings

```
<ddi:Fragment>
  <d:IfThenElse>
    <r:Agency>int.example</r:Agency>
    <r:ID>ifThen1</r:ID>
    <r:Version>1</r:Version>
    <r:InParameter>
      <r:Agency>int.example</r:Agency>
      <r:ID>inParam2</r:ID>
      <r:Version>1</r:Version>
      <r:Alias>genderOfRespondent</r:Alias>
      <r:CodeRepresentation>
        <r:CodeListReference>
          <r:Agency>int.example</r:Agency>
          <r:ID>codeListGender</r:ID>
          <r:Version>1</r:Version>
          <r:TypeOfObject>CodeList</
r:TypeOfObject>
        </r:CodeListReference>
      </r:CodeRepresentation>
    </r:InParameter>
    <d:IfCondition>
      <r:Command>
```

```
      <r:ProgramLanguage>MetaEnglish</
r:ProgramLanguage>
      <r:CommandContent>
genderOfRespondent == Female </
r:CommandContent>
    </r:Command>
  </d:IfCondition>
  <d:ThenConstructReference>
    <r:Agency>int.example</r:Agency>
    <r:ID>seq2</r:ID>
    <r:Version>1</r:Version>
    <r:TypeOfObject>Sequence</
r:TypeOfObject>
  </d:ThenConstructReference>
</d:IfThenElse>
</ddi:Fragment>
```



Question representations

- QuestionGrids
- QuestionBlocks
- Scale questions
- Ranked questions
- Point distribution questions



QuestionGrids

Generalized Interviewing Techniques FIELD 2005

Forms Answer Navigate Options Help

[F1]--Help
Current HH Member #1

I would like to begin the interview by getting a listing of all the people who live with you in this household -- just their first names, age, and relationship to you. Let's start with you.

What is your name?

- ♦ ENTER Name
- ♦ RECORD the selected respondent as the first person

| | Name | ReToMainR | Sex | Age | BirthYr | English |
|--------|----------|-----------|-----|-----|---------|---------|
| HHL[1] | JOHN DOE | 1 | 1 | 50 | 1964 | 1 |
| HHL[2] | | | | | | |
| HHL[3] | | | | | | |
| HHL[4] | | | | | | |
| HHL[5] | | | | | | |
| HHL[6] | | | | | | |

Name 001 3/26/2014 8:23:21 AM Version Date: 6/14/2005 Version Time: 7:39AM



QuestionGrids - 3.1

```
<d:QuestionItem isVersionable="true" id="Q29">
```

```
<d:QuestionItemName xml:lang="English">Name</d:QuestionItemName>
```

```
<d:QuestionText isStructureRequired="true" xml:lang="English">
```

```
<d:LiteralText>
```

```
<d:Text>What is your name?</d:Text>
```

```
</d:LiteralText>
```

```
</d:QuestionText>
```

```
<d:CodeDomain>
```

```
<r:CodeSchemeReference>
```

```
<r:ID/>
```

```
</r:CodeSchemeReference>
```

```
</d:CodeDomain>
```

```
</d:QuestionItem>
```

```
<l:Variable id="V29">
```

```
<l:VariableName>GIT.HHL.HHL[1].Name</l:VariableName>
```

```
<r:Label type="var">Name</r:Label>
```

```
<l:QuestionReference>
```

```
<r:ID>Q29</r:ID>
```

```
</l:QuestionReference>
```

```
<l:Representation>
```

```
<l:TextRepresentation maxLength="40"/>
```

```
</l:Representation>
```

```
</l:Variable>
```

```
<l:VariableGroup id="varGroup5">
```

```
<r:UserID
```

```
type="G">GIT.HHL.HHL[1]</r:UserID>
```

```
<l:GroupType>array</
```

```
l:GroupType>
```

```
<r:Label xml:lang="English"/>
```

```
<r:UniverseReference>
```

```
<r:ID>U20</r:ID>
```

```
</r:UniverseReference>
```

```
<l:VariableReference>
```

```
<r:ID>V28</r:ID>
```

```
</l:VariableReference>
```

```
<l:VariableReference>
```

```
<r:ID>V29</r:ID>
```

```
</l:VariableReference>
```

```
<l:VariableReference>
```

```
<r:ID>V31</r:ID>
```

```
</l:VariableReference>
```



QuestionGrids - 3.2

```
<ddi:Fragment>
  <d:QuestionGrid>
    <r:Agency>int.example</r:Agency>
    <r:ID>grid1</r:ID>
    <r:Version>1</r:Version>
    <d:GridDimension rank="1">
      <d:Roster baseCodeValue="1"
codelterationValue="1" minimumRequired="1"
maximumAllowed="25"
      > </d:Roster>
    </d:GridDimension>
    <d:StructuredMixedGridResponseDomain>
      <d:GridResponseDomain>
        <d:TextDomain>
          <r:Label>
            <r:Content>Name</r:Content>
          </r:Label>
        </d:TextDomain>
      </d:GridResponseDomain>
      <d:GridResponseDomain>
```

```
ddi:Fragment>
  <d:QuestionItem>
    <r:Agency>int.example</r:Agency>
    <r:ID>question1</r:ID>
    <r:Version>1</r:Version>
    <d:QuestionItemName>
      <r:String>q1</r:String>
    </d:QuestionItemName>
    <r:OutParameter>
      <r:Agency>int.example</r:Agency>
      <r:ID>outParam1</r:ID>
      <r:Version>1</r:Version>
      <r:Alias>firstName</r:Alias>
      <r:TextRepresentation/>
    </r:OutParameter>
    <d:QuestionText audienceLanguage="en-US">
      <d:LiteralText>
        <d:Text>What is your name?</d:Text>
      </d:LiteralText>
    </d:QuestionText>
    <d:TextDomain/>
  </d:QuestionItem>
</ddi:Fragment>
```

<



Scale questions - 3.0

On a scale of 0 to 100, what do you think are the chances that your income will keep up with inflation for the next five years?

Keep Up w/ Infl

```
<d:QuestionItem id="Q387">
  <d:QuestionItemName xml:lang="en-US">Keep Up w/
Infl</d:QuestionItemName>
  <d:QuestionText isStructureRequired="true"
xml:lang="English">
  <d:LiteralText>
    <d:Text>On a scale of 0 to 100, what do you think are
the chances that your income will keep up with inflation for the
next five years?</d:Text>
  </d:LiteralText>
</d:QuestionText>
<d:CodeDomain>
  <r:CodeSchemeReference>
    <r:ID/>
  </r:CodeSchemeReference>
</d:CodeDomain>
</d:QuestionItem>
```

```
...
<l:Variable id="V387">
  <l:VariableName>KeepUp</l:VariableName>
  <r:Label xml:lang="en-US">Scale Intro</r:Label>
  <r:Label type="var">KeepUp</r:Label>
  <l:QuestionReference>
    <r:ID>Q387</r:ID>
  </l:QuestionReference>
  <l:Representation>
    <l:NumericRepresentation type="Integer">
      <r:NumberRange>
        <r:Low>0</r:Low>
        <r:High>100</r:High>
      </r:NumberRange>
    </l:NumericRepresentation>
  </l:Representation>
</l:Variable>
```




Scale questions - 3.2

On a scale of 0 to 100, what do you think are the chances that your income will keep up with inflation for the next five years?

Keep Up w/ Infl

```
<ddi:Fragment>
```

```
<d:QuestionItem>
```

```
<r:Agency>int.example</r:Agency>
```

```
<r:ID>Q1</r:ID>
```

```
<r:Version>1</r:Version>
```

```
<d:QuestionItemName>
```

```
<r:String>KeepUp</r:String>
```

```
</d:QuestionItemName>
```

```
<d:QuestionText audienceLanguage="en-US">
```

```
<d:LiteralText>
```

```
<d:Text>On a scale of 0 to 100, what do you think are the
chances that your income will keep up with inflation for the
next five years?</d:Text>
```

```
</d:LiteralText>
```

```
</d:QuestionText>
```

```
<d:ScaleDomain>
```

```
<r:ScaleDimension>
```

```
<r:NumberRange>
```

```
<r:Low>0</r:Low>
```

```
<r:High>100</r:High>
```

```
</r:NumberRange>
```

```
<r:ValueIncrement startValue="0"
endValue="100" increment="1"/>
```

```
</r:ScaleDimension>
```

```
<r:Label>
```

```
<r:Content>Keep Up w/ Infl</r:Content>
```

```
</r:Label>
```

```
</d:ScaleDomain>
```

```
</d:QuestionItem>
```

```
</ddi:Fragment>
```



Ranked questions

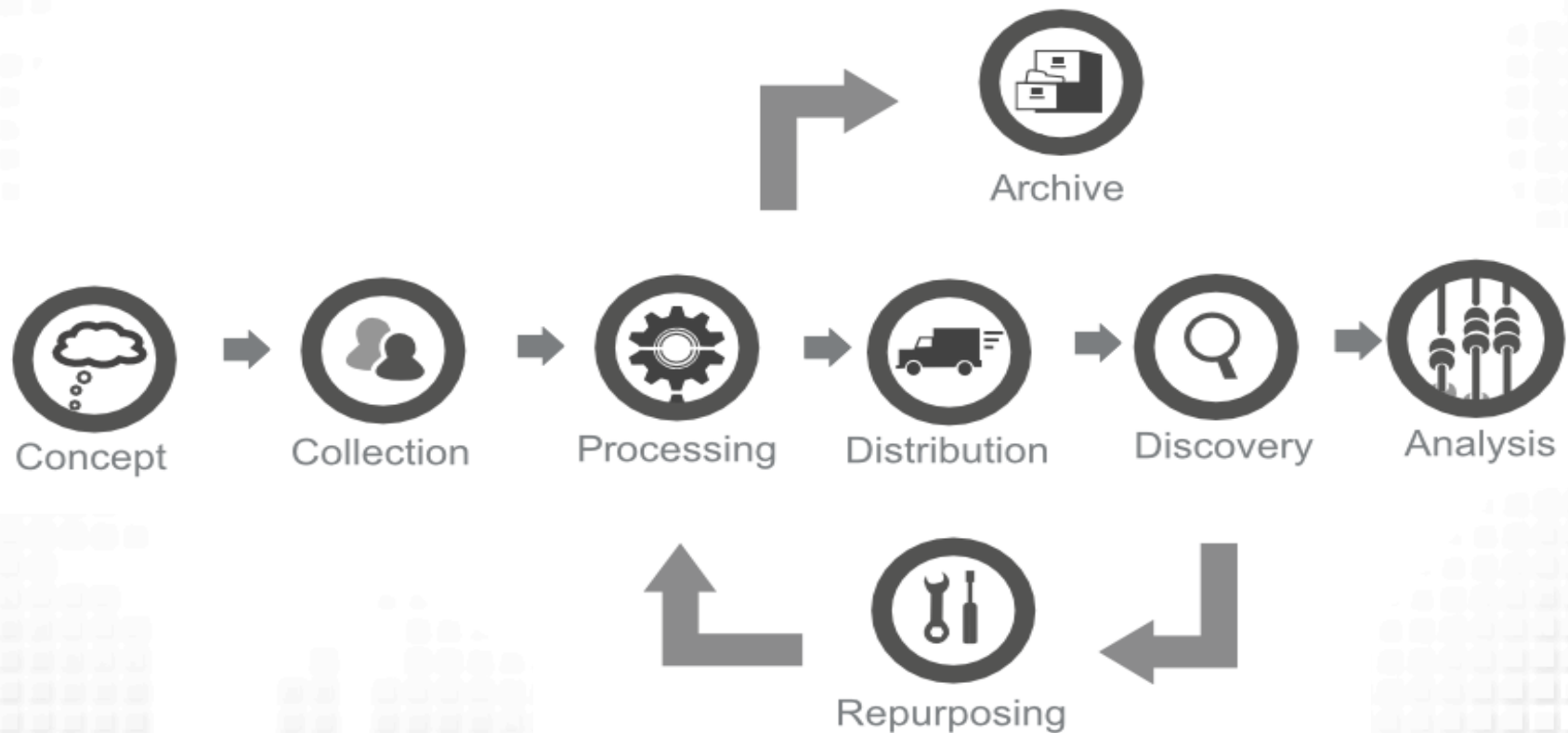
```
<ddi:Fragment>
  <d:QuestionGrid>
    <r:Agency>int.example</r:Agency>
    <r:ID>grid1</r:ID>
    <r:Version>1</r:Version>
    <d:GridDimension rank="1">
      <d:Roster baseCodeValue="1"
        codeIterationValue="1"
          minimumRequired="0"
          maximumAllowed="100"
        > </d:Roster>
    </d:GridDimension>
    <d:StructuredMixedGridResponseDomain>
      <d:GridResponseDomain>
        <d:TextDomain>
          <r:Label>
            <r:Content>Name</r:Content>
          </r:Label>
        </d:TextDomain>
      </d:GridResponseDomain>
    </d:StructuredMixedGridResponseDomain>
  </d:QuestionGrid>
</ddi:Fragment>
```

```
<d:GridResponseDomain>
  <d:RankingDomain>
    <r:RankingRange>
      <r:MinimumValue included="true">1</r:MinimumValue>
      <r:MaximumValue included="true">5</r:MaximumValue>
    </r:RankingRange>
  </d:RankingDomain>
</d:GridResponseDomain>
<d:GridResponseDomain>
  <d:DistributionDomain>
    <r:DistributionValue>100</r:DistributionValue>
  </d:DistributionDomain>
</d:GridResponseDomain>
</d:StructuredMixedGridResponseDomain>
</d:QuestionGrid>
</ddi:Fragment>
```



DDI 3.2 is much more powerful to

Document the full survey lifecycle in a standard manner





Next Step

- DDI3.2
- MQDS 5?
- Blaise 5?
- Colectica 5?

Thank YOU!