A Qualitative Data Model for DDI

Larry Hoyle and the DDI Qualitative Data Working Group: Louise Corti (Chair), Arofan Gregory (Convenor), Agustina Martinez (Co-Architect), Joachim Wackerow (Technical Implementation Committee representative), Eirik Alvar, Noemi Betancort Cabrera, Damien Gallagher, Tobias Gebel, Jani Hautamaki, Arja Kuula, Steve McEachern, Cornelia Zuell

Abstract

The Qualitative Data Model Working Group was established in January 2010 with the charge "To develop a robust XML-based schema for qualitative data exchange (compliant with DDI) and encourage tools development based upon these needs." This report describes the preliminary model developed by that group via online meetings, and working meetings in Gothenburg (2011) and Bergen (2012).

This model, described in UML, was developed to cover three main scenarios:

1. Qualitative data collections needing metadata at the object level only
2. Qualitative data collections where segments of objects need to be delineated and described and where segments of different physical representations of the same logical objects possibly need to be linked
3. Qualitative data collections as in the second case where related quantitative data have been generated through techniques such as text mining

Introduction

The working group began in April of 2010 by gathering a set of use cases and summarizing them in terms of types of qualitative data and types of metadata elements potentially needed. Analysis of the use cases led to enumerating the three main scenarios described above.

At the Gothenburg meeting the group developed a first draft of a model expressed in UML. This model was refined further in the Bergen meeting and is presented here.

Examples of Qualitative Data

The term "qualitative data" applies to a wide range of collections of digital, and, potentially real, objects. Qualitative data analysis tools can work with text, audio and video recordings, and still images. In developing the metadata model we considered cases where identification of segments of original real objects might need to be linked to some sort of digital representation of the objects. An example might be a collection of preserved archeological artifacts of which multiple digital images were taken. We did not, however, describe any methods by which segments of arbitrary physical objects might be identified. We will need to explore whether others have addressed this issue. One place to start might be techniques used by game hardware to recognize facial features, hands, feet, etc., from video input.

Use Cases

Object Level Only

Both the UK Data Archive and the Finnish Social Science Data Archive offered use cases in which a vector of arbitrary attributes might be used to select a subset of qualitative data objects from an archive. Examples might be selecting all of the audio and transcripts of interviews where an interviewee was a female over 40 years of age, or all of the video clips where the season was winter.

These use cases highlight the need for a flexible system of assigning arbitrary attributes to digital objects. Allowing for using these attributes by reference from a repository will make comparison or aggregation across collections possible.

Describing Segments

One use case from the Data and Networked Services (DANS) requiring the delineation and description of segments involved a collection of video clips which had also been transcribed (possibly into multiple languages). A Web application allowed for searching for segments of the clips by attributes. Both the video and its transcription were then displayed to the searcher.

From this use case it became clear that some object in the model was needed to represent the logical content of the segment which was expressed in several forms (the video and the transcript).

Another use case in this category is that of Computer Assisted Qualitative Data Analysis (CAQDAS). In this scenario, a researcher identifies segments of the original objects and categorizes them, possibly assigning codes or memos to them.

Segments with Quantitative Data

The most complex use case presented involved some sort of automated analytical process which assigned quantitative variables to segments of qualitative objects. One specific example is that of text mining, where the transcripts of interviews are divided into segments (e.g., sentences or paragraphs). These segments are then processed by statistical procedures that assign them values on some calculated dimensions (variables) or assign them to categorical (grouping) variables.

The set of quantitative variables generated for each segment (a logical record in DDI terms) forms another representation of the quantitative logical object (the segment) and can therefore be linked to all of the other physical representations – for example, a video segment from which the transcript was taken.

Modeling Tool

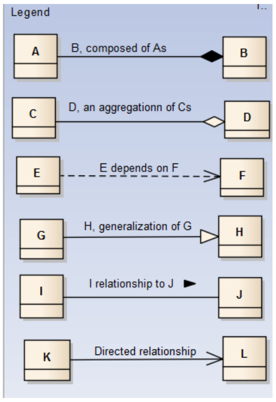
At the Gothenburg meeting we encountered some problems with exchange of UML models between different software packages. In order to allow exchange of the UML model among the group, for the Bergen meeting we decided on Enterprise Architect as the common software package. The model described here is available as an Enterprise Architect (eap) file. The class diagrams are also available as pdf files, including a text document describing the attributes and relationships of the objects in the model.

The Model

Small Subset of UML Modeling Elements

The model was developed with a small subset of all of the UML modeling elements:

- Composition – the parent is composed of a set of objects. If one of those objects ceases to exist, the parent object ceases to exist
- Aggregation – the parent is associated with a collection of objects. If one of those objects ceases to exist, the parent still exists
- Dependency – one object depends on another in some way
- Generalization - a parent object is a more general version of a child element. The child inherits properties from the parent, but may be more specialized having additional properties
- Relationship – one object is related to another in some fashion



- Directed Relationship – the relationship between two objects is one way, for example, A might send something to B

The Model Elements

The model consists of a number of elements and their relationships. The elements also have properties. A list of the elements, their definitions, and a list of their properties follows.

Collection – An organized set of "qualitative" (non-quantitative) data objects that derive from a particular piece of research or fieldwork method (as opposed to quantitative/numeric "datasets" described in DDI). This might typically be a subcollection of a larger DDI study, or a whole DDI study. These data objects are primary data collected/created by the research process.
Examples/Comments: This is not necessarily a DDI "study" along with context (related materials). A list of files, a grouping put together for a research purpose.
Properties: Name, Label, Description, Creation Date, Citation, Temporal Coverage, Spatial Coverage, Topical Coverage, Identifiable, Versionable, Maintainable

Logical Resource - A "Data" object. Logical unit of a data collection distinct from the physical storage/files.
Examples/Comments: an interview, a picture, an observation.
Properties: Name, Label, Description, Type, IsOriginal, Digital Version Available, Creation Date, Citation, Temporal Coverage, Spatial Coverage, Topical Coverage, Identifiable, Versionable

Event - Any activity that produces or influences the data, transforms data or processes data to create information (coding). Events are time-bound and typically have a setting/context.
Examples/Comments: a data collection/gathering activity, an ingest/processing event. Could overlap with DDI. Can be a change event. Could include activities that impact on any aspect of the research. An interview event, a focus group, transcribing, coding, processing/indexing.
Properties: Name, Label, Description, Type, Event Date/Time, Event location, Event Setting, Identifiable, Versionable, Maintainable

Entity - A thing associated with an event or a logical resource. Synonym = agent.
Examples/Comments: data collection manager, interviewee/participant, third party, interviewer, software agent, analytic agents: e.g., researchers, automated. Not a funding agency, but directly related.
Properties: Name, Label, Description, Identifiable, Versionable, Maintainable

Function - The type of activity an entity plays in relation to an event or a resource.
Examples/Comments: interviewer, interviewee, author, etc.
Properties: Name, Label, Description, Type, Identifiable, Versionable, Maintainable

Effect - The impact of an event on a logical resource.
Examples/Comments: A financial crash occurring during interviews about work and employment; a person interrupting an interview.
Properties: Name, Label, Description, Type, Identifiable, Versionable, Maintainable

Attribute – A configurable property about events, actors, and logical resources and collections. This is one type of descriptive metadata.
Examples/Comments: A characteristic; e.g., age of an interviewee (actor), or time period or geography. Can be a variable (attribute declaration) but within a narrowed universe (a sub class).
Properties: Name, Label, Description, Data Type, Value, Is Value Required, Type, Identifiable, Versionable, Maintainable

Descriptive Metadata (Abstract) - Metadata used for managing and discovering objects within a collection.
Examples/Comments: attributes in data lists/finding aids, creator and title, classification or subject terms, sampling characteristics. Not necessarily analytical metadata, but could be. Comment from the PREMIS Data Dictionary for Preservation Metadata: "Descriptive metadata can be important both for discovery of archived

resources and during preservation planning" (see http://www.loc.gov/standards/premis/v2/premis-dd-2-2.pdf).
Properties: No properties are defined for this abstract object.

Research Metadata (Abstract) **-** Organizational categories, analytic codes, memos used in analysis - create and use.
Examples/Comments: metadata about segments of an interview.
Properties: No properties are defined for this abstract object.

Approach (Abstract)- The process for accomplishing something.

Method (Abstract) - Approach to data collection, analysis and processing.

Dataset – (Quantified Dataset) As in DDI. A quantification derived from logical resources.
Examples/Comments: Text mining output. A coded numeric dataset. (See PhysicalInstance in DDI).
Notes: We may need some way for a LogicalRecord in a DDI dataset to indicate the variable which links back to the Physical Segment from which the LogicalRecord was derived.

Logical Segment - All or part of a logical resource. The 'thing' that is referenced for the purposes of attaching analytic metadata.
Examples/Comments: A textual 'extract', audio/video clip, area in an image.
Properties: Name, Label, Description, Type, Identifiable, Versionable, Maintainable

Physical Segment (Abstract) The physical encoding of the logical segment, made up of information allowing location of the segment within the associated physical instance.
Examples/Comments: e.g., characters 30-40 of the third line of a textual transcript of an interview; X and Y axis coordinates for a polygon containing the face of a person in an image. Other examples of objects for which segments might be described and how that segment might be described include:

XML: a selection described by an Xpath - node(s)
TIME-BASED: objects with a timeline - a format, start/end time
IMAGE: an area – format, shape and coordinates (x,y)
RELATIONAL DATABASES: tables, col, row; SQL query, DB types, SQL type
WEBPAGE: URLs, Xpath
TRIPLE STORE: graph, end point, type of database, SPARQLquery
3D IMAGE: (to be determined) (an example might be a bounding parallelepiped)

Properties: No properties are defined for this abstract object.

Physical Storage Instance - Encoding of a logical resource in one or more files or data sources.
Examples/Comments: Image files, XML docs, databases, AV files, AV streams. Use standard DDI file properties.
Properties: Name, Label, Type, Location, IsDigital, Data Fingerprint, Creation Date, File identifier, Identifiable, Versionable, Maintainable

File Structure - The information describing the format of a file or similar data storage instance.
Examples/Comments: an XML schema, the description of a relational database table structure. Mime type for simple file formats.
Properties: Name, Label, Description, Specification Location, TypeSystem, TypeValue, Link To External Documentation, Identifiable, Versionable, Maintainable

Structure Schema - a description of how to locate the segment in the physical storage instance associated with the segment. A referencing method.
Examples/Comments: Drawing a polygon around the face of a person in an image; identifying the last 30 seconds of an audio clip; line and character offsets in a textual interview.
Properties: Name, Label, Description, Type, Identifiable, Versionable, Maintainable

Analytic Category - Concept used to organize/classify objects for the purposes of management and filtering during analysis, typically expressed as an alphanumeric string. Can be hierarchical.
Examples/Comments: gender, age. These may often be used to populate attributes.
Properties: TermDescription, Value, Type, Identifiable, Versionable, Maintainable

Analytic Code - A short alphanumeric string used to classify segment for the purposes of analysis.

Examples/Comments: concept applied to a segment. Can be hierarchical. Work: full-time and part-time.

Properties: TermDescription, Value, Type, Identifiable, Versionable, Maintainable

Memo - An annotation used for the purposes of fieldwork or analysis. Can apply to segments, analytic categories and analytic codes.

Properties: Name, Label, Description, Data type, Type, Creation Date, Content, Link to External Content, Identifiable, Versionable, Maintainable

Methods Objects

Instrument (Abstract) - Instruction or protocol for data collection, analysis and processing.

Tool (Abstract) - Mechanism /equipment for data collection, analysis and processing.

Settings - The configuration of the tool.

Examples/Comments: Parameters for a software program, settings on dials of an instrument.

Properties: Name, Label, Description, Value, Type, Identifiable, Versionable, Maintainable

Data Collection Tool - Mechanism /equipment by which data are captured and or recorded.

Examples/Comments: pen and paper; Microsoft software, BLAISE computer-assisted Interview software, video camera, recording device, simulation software, brain scanning.

Properties: Name, Label, Description, Model/Version, Manufacturer, Type, Identifiable, Versionable, Maintainable

Analysis Tool - Mechanism /equipment by which data are analyzed.

Examples/Comments: NVivo, pen and paper.

Properties: Name, Label, Description, Model/Version, Manufacturer, Type, Identifiable, Versionable, Maintainable

Processing Tool - Mechanism /equipment by which data are processed.

Examples/Comments: SPSS. MS Word. XML editors.

Properties: Name, Label, Description, Model/Version, Manufacturer, Type, Identifiable, Versionable, Maintainable

Data Collection Method - Approach to gathering data.

Examples/Comments: Sampling, fieldwork, observation, etc.. Note that other metadata standards, e.g., PREMIS (Preservation Metadata Maintenance Activity), may offer the capability of documenting data collection method.

Properties: Name, Label, Description, Type, Identifiable, Versionable, Maintainable

Analysis Method - Approach to data reduction/coding.

Examples/Comments: grounded theory; approaches used for coding interview text; data mining.

Properties: Name, Label, Description, Type, Identifiable, Versionable, Maintainable

Processing Method - Approach to processing or transforming data.

Examples/Comments: Transcription of an audio recording or historical document; lemmatization of text; validation; anonymization; reformatting of a document, XML e.g., TEI mark-up.

Properties: Name, Label, Description, Type, Identifiable, Versionable, Maintainable

Data Collection Instrument - Instruction or protocol for gathering data. This is not a tool.

Examples/Comments: Topic guide. Questions. Template. Prompted tasks, e.g., story writing. Observation template. Transcription template.

Properties: Name, Label, Description, Type, Citation, Link to External Resources, Identifiable, Versionable, Maintainable

Analysis Instrument - Instruction or protocol for analysis.

Examples/Comments: instructions for coding.

Properties: Name, Label, Description, Type, Citation, Link to External Resources, Identifiable, Versionable, Maintainable

Processing Instrument - Instruction or protocol for processing.

Examples/Comments: Lemmatization rules, transcription instructions/templates. Anonymization rules.

Properties: Name, Label, Description, Type, Citation, Link to External Resources, Identifiable, Versionable, Maintainable

Data Collection Mode - The technique by which data are captured.

Examples/Comments: Based on the experience of the object of study, how a respondent interacts with the interviewer - may produce mode effects: e.g., face to face, computer assisted personal interviewing (CAPI), paper and pencil interviewing (PAPI), computer assisted telephone interviewing (CATI); self-administered questionnaires (SAQ), Photographing.

Properties: Name, Label, Description, Type, Identifiable, Versionable, Maintainable

Physical Segment Objects

SegmentOfText - A description of the location of a subset of a text document. A subclass of a Physical Segment of type text.

Examples/Comments: A section of text selected in a qualitative data analysis program for coding. The response to a question in the transcript of an interview. A paragraph or sentence.

Properties: Name, Label, Description, Creation Date, Identifiable, Versionable, Maintainable

LineParameter - SegmentOfText parameters based on lines and characters.

Examples/Comments: A partial segment starting at character 3 of line 5 and ending with character 7 of line 10.

Properties: Name, Label, Description, Creation Date, StartLine, StartOffset, EndLine, EndOffset, Identifiable, Versionable, Maintainable

CharacterParameter - SegmentOfText parameters based on characters.

Examples/Comments: A partial segment starting at character 3 and ending with character 128346 of a document.

Properties: Name, Label, Description, Creation Date, StartCharOffset, EndCharOffset, Identifiable, Versionable, Maintainable

SegmentOfXML - A description of the location of a subset of an XML document. A subclass of a Physical Segment relating to XML documents.

Examples/Comments: Several row nodes from a table node from an XHTML document.t

Properties: Name, Label, Description, Creation Date, Identifiable, Versionable, Maintainable

SegmentOfWebPage - A description of the location of part of a Web page. A subclass of a Physical Segment regarding a Web page.

Examples/Comments: All of the top-level headers (H1) from an HTML document. Described using XPointer (XPath) which should work for well-formed HTML, but may not work in some non-well-formed cases.

Properties: Name, Label, Description, Creation Date, URL, Identifiable, Versionable, Maintainable

Node - SegmentOfXML parameters based on an Xpointer node.

Examples/Comments: An XPointer to a row node from a table node from an XHTML document.

Properties: Name, Label, Description, Creation Date, xPtrExp, Identifiable, Versionable, Maintainable

SegmentOfAudio - A description of the location of part of an audio file.

A subclass of a Physical Segment regarding audio.

Examples/Comments: An audio clip beginning at 4.54 seconds into the full audio sample and ending at 9.375 seconds. Enumerated clipTypes for Audio are:

- “byte”
- “smil”
- “midi”
- “time”
- “tcf”

- “other”

Properties: Name, Label, Description, Creation Date, Identifiable, Versionable, Maintainable

SegmentOfVideo - Definition: A description of the location of part of a video recording.

A subclass of a Physical Segment regarding video

Examples/Comments: A clipType of “time”, begin value of 4.54, and end value of 9.375 defining a video clip beginning at 4.54 seconds into the full video sample and ending at 9.375 seconds. Enumerated clipTypes for Video are

- “byte”
- “smil”
- “midi”
- “smpte-25”
- “smpte-24”
- “smpte-df30”
- “smpte-ndf30”
- “smpte-df29.97”
- “smpte- ndf29.97”
- “time”
- “tcf”
- “other”

Properties: Name, Label, Description, Creation Date, Identifiable, Versionable, Maintainable

BeginEnd - Parameters based on a description of a beginning and an ending. A clipType value from a controlled vocabulary defines the units of measurement for the beginning and ending.

Examples/Comments: A clipType of “time”, begin value of 4.54, and end value of 9.375 defining an audio clip beginning at 4.54 seconds into the full audio sample and ending at 9.375 seconds.

Properties: Name, Label, Description, Creation Date, ClipType, OtherClipType, ClipBegin, ClipEnd, Identifiable, Versionable, Maintainable

SegmentOfImage - A description of the location of part of an image. A subclass of a Physical Segment regarding images.

Properties: Name, Label, Description, Creation Date, Identifiable, Versionable, Maintainable

Area - Parameters based on a description of an area. A shape value from a controlled vocabulary specifies the meaning of a comma-separated list of coordinates.

Examples/Comments: The definition of a rectangular section of a digital photograph using the four pixel values left-x, top-y, right-x, bottom-y.

Properties: Name, Label, Description, Creation Date, Shape, Coords, Identifiable, Versionable, Maintainable

SegmentOfTripleStore - A description of the location of a subset of a Triple Store. A subclass of a Physical Segment regarding a Triple Store.

Examples/Comments: An extract returned from a triple store as the result of a SPARQL query.

Properties: Name, Label, Description, Creation Date, EndPoint, SPARQL Query, Graph, Identifiable, Versionable, Maintainable

SegmentOfRelationalDatabase - A description of the location of an extract from a relational database. A subclass of a Physical Segment regarding a relational database.

Examples/Comments: Results of a query joining two tables from a relational database.

Properties: Name, Label, Description, Creation Date, Identifiable, Versionable, Maintainable

SQLSelection - An SQL query which defines a set of rows (a table) selected from an SQL capable relational database

Examples/Comments: The description of an extract from a relational database defined by the query:

select * from a,b where a.id=b.id and gender=“F”.

If DatabaseType and Database version are not present the SQL query is assumed to be “standard” SQL.

Properties: Name, Label, Description, Creation Date, SQL Statement, DatabaseType, DatabaseVersion, SchemaName, Identifiable, Versionable, Maintainable

SegmentOfDDIDescribedData - A description of the location of part of data described by a DDI Instance. A subclass of a Physical Segment regarding data described by a DDI Instance.
Properties: Name, Label, Description, Creation Date, Identifiable, Versionable, Maintainable

NOTE: We could also have a SegmentWithDimensionalIndices (SegmentOfMultidimensional) that could be used to describe any piece of a multidimensional structure -- for example, a spreadsheet. It could have a subclass (BoundingCube) describing a hypercube subset with a parameter for number of dimensions and a parameter with the tuple for the "lower left" corner and the tuple for the "upper right" corner. The subclass would be repeatable. This would allow the description of multiple multi-dimensional bounding boxes -- e.g., two tables from a single sheet, or all cells in the upper left corner of several sheets in a spreadsheet file.

Essentially this is multiple bounding hyper-cubes. There could also be a class allowing for naming of the dimensions.

SegmentOfMultidimensional - A description of the location of part of a data object using multiple dimensions.
Examples/Comments: A simple example would be a segment of a spreadsheet file where the dimensions are "row", "column", and "sheet". The segment in this case would be a bounding box cube described by two triples, for example, (4,5,6), (10,11,12) where row 4, column 5, and sheet 6 represent the low corner of the cube and (10,11,12) represents the high corner of the cube.
Dimensions are named, as with "row", "column", and "sheet" in the previous example. Levels of a dimension may also be named. For spreadsheet columns names might be "A", "B", etc.
This is a generic method of description that may apply to a variety of object types. A religious text, for example might have dimensions of "book", "chapter", and "verse". A film might have dimensions of "frame", "horizontal", "vertical". The latter might allow for the description of a clip of the location of a character in the film across time.
Properties: Name, Label, Description, Creation Date, NumberOfDimensions, DimensionNames, Identifiable, Versionable, Maintainable

BoundingCube - A BoundingCube contains the two tuples describing the location of a sub hypercube of the data object.
Examples/Comments: A simple example would be a segment of a spreadsheet file where the dimensions are "row", "column", and "sheet". The segment in this case would be a bounding box cube described by two triples, for example, (4,5,6), (10,11,12) where row 4, column 5, and sheet 6 represent the low corner of the cube and (10,11,12) represents the high corner of the cube.
Dimensions are named, as with "row", "column", and "sheet" in the previous example. Levels of a dimension may also be named. For spreadsheet columns names might be "A", "B", etc.
This is a generic method of description that may apply to a variety of object types. A religious text, for example might have dimensions of "book", "chapter", and "verse". A film might have dimensions of "frame", "horizontal", "vertical". The latter might allow for the description of a clip of the location of a character in the film across time.
Properties: Name, Label, Description, Creation Date, LowPoint, HighPoint, Identifiable, Versionable, Maintainable

DimensionNames - DimensionNames lists the names of the levels of a single dimension.
Examples/Comments: An example might be labels of spreadsheet columns of "A", "B", etc., or the "book" dimension of a religious text might have labels "Genesis, Exodus"….
Properties: Name, Label, Description, Creation Date, DimensionNumber, LabelList, Identifiable, Versionable, Maintainable

Glossary of Properties

A list of the attributes used by objects follows. Each attribute is preceded by the number of objects to which it applies. For attributes used by only a few objects, a list of those objects follows the definition.

5 Citation - Provides bibliographic citation information

1 ClipBegin - A string, appropriate to the @clipType, describing the start of the clip (BeginEnd)

1 ClipEnd - A string, appropriate to the @clipType, describing the end of the clip (BeginEnd)

1 ClipType - The annotation method used to describe the audio or video clip (BeginEnd)

1 Content - The text associated with the object (Memo)

1 Coords - A comma-delimited string of coordinates (Area)

21 Creation Date - Date of the creation of the resource (e.g., a transcription)

1 Data Fingerprint – An identifier for the data object, derived from its content, e.g., checksum (Physical Storage Instance)

2 Data Type - (e.g., string, date, time, URL, …) (Attribute, Memo)

1 DatabaseType - the type of the database, e.g., "Oracle" (SQLSelection)

1 DatabaseVersion - The version of the database, e.g., 10.0 (SQLSelection)

39 Description - provides additional detailed information regarding the object

1 Digital Version Available - Was it ever not digital? Born digital? Boolean: (Logical Resource)

1 DimensionNames - A comma-delimited list of quoted strings listing the names of the dimensions. Example from above: ("row", "column", "sheet") (SegmentOfMultidimensional)

1 DimensionNumber - The number of the dimension having labels, for example, the columns dimension of a spreadsheet might appear second in the tuples describing a location. In this case the value might be "2". (DimensionNames)

1 Dimensions - (Physical Storage Instance)

1 EndCharOffset – Identifies the end of a segment by specifying the number of characters from the start of the document (CharacterParameter)

1 EndLine - Identifies the end of a segment by specifying the number of lines from beginning of the document (LineParameter)

1 EndOffset - Identifies the end of a segment by specifying the number of characters from start of the appropriate line (LineParameter)

1 EndPoint - The EndPoint(s) used for the extract (SegmentOfTriplestore)

1 Event Date/Time - Start and end attributes (Event)

1 Event Location – Geography (Event)

1 Event Setting - environment: contextual, summary, thumbnail (Event)

1 File Identifier - Used for resolving, e.g., a URI or DOI (Physical Storage Instance)

1 Graph - URI of the graph to which the SPARQLquery applies. If not specified, either the query applies to all the graphs available from the Endpoint or it is specified directly on the query. (SegmentOfTripleStore)

1 HighPoint - The tuple listing the high values on each dimension (BoundingCube)

42 Identifiable - Can the object be assigned an id? For objects which can be referenced. Boolean

1 Is Value Required - Is a Value attribute required? Boolean (Attribute)

1 IsDigital - True if the object is in digital form. Boolean (Physical Storage Instance)

1 IsOriginal - Is this the original source? Boolean (Logical Resource)

40 Label - A structured display label for the element. Label provides display content of a fully human-readable display for the identification of the element.

1 LabelList - A quoted comma-separated string containing the dimension labels. Example: ("Genesis", "Exodus") (DimensionNames)

1 Link To External Documentation - A reference to an external source of information (File Structure)

1 Link to External Content – A reference to an external document, e.g., a Web page (Memo)

3 Link to External Resources – A reference to the location of external resources for the instrument. These could include material to be displayed, or some Web-based services (Data Collection Instrument, Analysis Instrument, Processing Instrument)

1 Location – Where the Instance is to be found (Physical Storage Instance)

1 LowPoint - The tuple listing the low values on each dimension (BoundingCube)

41 Maintainable – Is the object one that can be maintained over time by a responsible party? Specification of the responsible party is required. Boolean

3 Manufacturer – The creator of the tool (Data Collection Tool, Analysis Tool, Processing Tool)

3 Model/Version – A model identifier or number and additionally a version number (Data Collection Tool, Analysis Tool, Processing Tool)

40 Name – A human-readable name given the object being identified

1 NumberOfDimensions - The number of dimensions used to describe the segment. A cell in a multi-column spreadsheet might be identified by three dimensions – row number, column number, sheet number (SegmentOfMultidimensional)

1 OtherClipType - A text description recommended if @clipType = “other” (BeginEnd)

1 SPARQLquery - The SPARQLquery string to be executed (SegmentOfTriplestore)

1 SQLStatement - The SQL query string (SQLselection)

1 SchemaName - The name of the database schema being queried (SQLselection)

1 Shape - The shape of the segment, one of the following

“Rect”	a rectangular area
“Circ”	a circular area
“poly”	an irregular polygon area

The @shape should be used as in HTML 4.
Reference: http://www.w3.org/TR/html401/struct/objects.html (Area)

2 Spatial Coverage - Description of the geographic coverage of the data (Collection, Logical Resource)

1 Specification Location - The location of a formal specification for the type of file structure, e.g., DDI or TEI schema, relational database schema location, URI: (File Structure)

1 StartCharOffset - Identifies the start of a segment by specifying the number of characters from the start of the document (CharacterParameter)

1 StartLine - Identifies the start of a segment by specifying the number of lines from beginning of the document (LineParameter)

1 StartOffset - Identifies the start of a segment by specifying the number of characters from start of the appropriate line (LineParameter)

2 Temporal Coverage - A description of the time period(s) related to the data (Collection, Logical Resource)

1 TermDescription - Human-readable description; string: Optional. Field only to be used where meaning is not available elsewhere (Analytic Category, Analytic Code)

2 Topical Coverage - A description of the conceptual topics related to the data (Collection, Logical Resource)

22 Type - A description of the type of object. This should be selected from a controlled vocabulary.

1 TypeSystem – The system used to describe the type of file structure. The TypeValue will be interpreted in this context, e.g., mime type. This should be selected from a controlled vocabulary. (File Structure)

1 TypeValue – A string indicating the type of file structure, interpreted within the system indicated by TypeSystem, e.g., the mime type of text/html. This should be selected from the controlled vocabulary described by TypeSystem.

1 URL – The Uniform Resource Locator of the Web page (SegmentOfWebPage)

4 Value – This property contains the datum for the object. For categories and codes, the value should be selected from a controlled vocabulary (Analytic Category, Analytic Code, Settings, Attribute)

42 Versionable – An object for which, when once "published" (in the DDI sense), changes need to be tracked by a change in version.

1 xPtrExp - The Xpointer expression describing the xml range (Node)

Relationships

Each of the following relationship descriptions is listed in triples form with the name of the relationship listed in bold and the objects of the relationship in italics. A definition of the relationship follows the triple. Relationships are listed for the source of a relationship (a source relates to a target).

Collection

- One *Collection* **is composed of** 0 or more *Collection* - The Collection is composed Collections.
- One *Collection* **is composed of** 0 or more *Event* - The Collection is composed of Events.
- One *Collection* **is composed of** 0 or more *Attribute* - The Collection is composed of Attributes.
- One *Collection* **is composed of** 1 or more *Logical Resource* - The Collection is composed of at least one Logical Resource.
- One *Collection* **is composed of** 0 or more *Entity* - The Collection is composed of these objects.
- Zero or more *Collection* **uses** 0 or more *Tool* - The Collection uses a Tool. *(relationship)*
- One Collection **uses** 0 or more *Method* - The Collection uses a Method. *(relationship)*

Logical Resource

- Zero or more *Logical Resource* **results from** 0 or 1 *Method* - a Logical Resource results from a Method. *(dependency)*
- Zero or more *Logical Resource* **uses** 0 or more *Instrument* - a Logical Resource uses an Instrument. *(relationship)*
- One *Logical Resource* **is composed of** 0 or more *Logical Segment* - Logical Resource is composed of Logical Segments
- One *Logical Resource* **is composed of** 0 or more *Attribute*- Logical Resource is composed of attributes

Event

- One *Event* **is composed of** 0 or more *Attribute* - The Event is composed of these objects
- One or more *Event* **has** 0 or more *Effect* – A combination of Event has a combination of Effect *(relationship)*
- One or more *Event* **involves** 0 or more *Method* - The Event involves a Method *(dependency)*

Entity

- Zero or more *Entity* **has** 0 or more *Function* - An Entity has a Function *(undirected relationship see below)*
- One *Entity* **is composed of** 0 or more *Attribute* - The Entity is composed of these objects
- Zero or more *Entity* **appears in** 0 or more *Logical Resource* - An Entity appears in a Logical Resource *(relationship)*
- Zero or more *Entity* **uses** 0 or more *Method* - An Entity uses a Method *(relationship)*

Function

- Zero or more *Function* **is related to** 0 or more *Event* – A Function is related to an Event *(relationship)*
- Zero or more *Function* **is related to** 0 or more *Logical Resource* - An Function is related to a Logical Resource *(relationship)*
- Zero or more *Function* **has** 0 or more *Entity* - A Function has an Entity *(undirected relationship see above)*

Effect

- Zero or more *Effect* **on** 0 or more *Logical Resource* - A Effect may act on a Logical Resource *(relationship)*

Attribute

- *Attribute* **extends** *Descriptive Metadata* - An Attribute is a specific extension of Descriptive metadata

Descriptive Metadata (abstract) (not the source for any relationship)

Research Metadata

- Zero or more *Research Metadata* **results from** 0 to 1 *Event* - Research Metadata results from an Event *(dependency)*
- Zero or more *Research Metadata* Created By 0 to 1 *Entity* - Research Metadata is created by an Entity *(dependency)*
- Approach (abstract)

Method (abstract)

- *Method* **extends** *Approach* - A Method extends an Approach
- Zero or more *Method* **uses** 0 or more *Logical Resource* - A Method uses a Logical Resource *(dependency)*
- Zero or more *Method* **uses** 0 or more *Descriptive Metadata* - A Method uses Descriptive Metadata *(dependency)*
- Zero or more *Method* **uses** 0 or more *Research Metadata* - A Method uses Research Metadata *(dependency)*

Dataset (Quantified Dataset) (not the source for any relationship)

Logical Segment

- Zero or more *Logical Segment* **results from** 1 or more *Method* - A Logical Segment results from a Method *(dependency)*

Physical Segment

- Zero or more *Physical Segment* **part of** 1 or more *Physical Storage Instance* - A Physical Segment is part of a Physical Storage Instance *(dependency)*
- Zero or more *Physical Segment* **uses** 1 *Structure Schema* - A Physical Segment uses a Structure Schema *(dependency)*
- One or more *Physical Segment* **encodes** 1 or more *Logical Segment* - A Physical Segment encodes a Logical Segment *(relationship)*

Physical Storage Instance

- One or more *Physical Storage Instance* **uses** 0 to 1 *File Structure* - One or more Physical Storage Instances may use a File Structure *(relationship)*
- Zero or more *Physical Storage Instance* **results from** 0 or more *Tool* - A Physical Storage Instance may result from a Tool *(dependency)*
- *One or more Physical Storage Instance* **encodes one or more** *Logical Resource* - One or more Physical Storage Instances encode(s) one or more Logical Resource(s) *(aggregation)*

File Structure (not the source for any relationship)

Structure Schema (not the source for any relationship)

Analytic Category

- Zero or more *Analytic Category* **child of** 0 or more *Analytic Category* - An Analytic Category can be a child of an Analytic Category (hierarchical Categories) *(relationship)*
- Zero or more *Analytic Category* **attached to** 0 or more *Physical Segment* - An Analytic Category may be attached to Physical Segment(s) *(relationship)*
- Zero or more *Analytic Category* **attached to** 0 or more *Analytic Code* - An Analytic Category may be attached to Analytic Code(s) *(relationship)*
- Zero or more *Analytic Category* **attached to** 0 or more *Memo* - An Analytic Category may be attached to Memo(s) *(relationship)*
- *Analytic Category* **extends** *Research Metadata* - An Analytic Category is an extension of Research Metadata

Analytic Code

- Zero or more *Analytic Code* **child of** 0 or more *Analytic Code* - An Analytic Code can be a child of an Analytic Code (hierarchical Codes) *(relationship)*
- Zero or more *Analytic Code* **attached to** 1 or more *Physical Segment* - An Analytic Code may be attached to Physical Segment(s) *(relationship)*
- Zero or more *Analytic Code* **attached to** 0 or more *Analytic Category* - An Analytic Code may be attached to Analytic Category(s) *(relationship)*
- Zero or more *Analytic Code* **attached to** 0 or more *Memo* - An Analytic Code may be attached to Memo(s) *(relationship)*
- *Analytic Code* **extends** *Research Metadata* - An Analytic Code is an extension of Research Metadata

Memo

- Zero or more *Memo* **attached to** 1 or more *Physical Segment* - A Memo may be attached to Physical Segment(s) *(relationship)*
- Zero or more *Memo* **attached to** 0 or more *Analytic Category* - A Memo may be attached to Analytic Category(s) *(relationship)*
- Zero or more *Memo* **attached to** 0 or more *Analytic Code* - A Memo may be attached to Analytic Code(s) *(relationship)*
- *Memo* **extends** *Research Metadata* - A Memo is an extension of Research Metadata

Instrument

- Zero or more *Instrument* **implements** 0 or more *Method* - An Instrument implements a Method *(relationship)*

Tool

- Zero or more *Tool* **applies** 1 or more *Instrument* - A Tool Applies an Instrument *(relationship)*
- One *Tool* **has** 1 or more *Settings* - A Tool has Settings *(relationship)*

Settings (not the source for any relationship)

Data Collection Tool

- *Data Collection Tool* **extends** *Tool* - A Data Collection Tool extends Tool
- Zero or more *Data Collection Tool* **applies** 1 or more *Instrument* - A Data Collection Tool Applies an Instrument *(relationship)*
- Zero or more *Data Collection Tool* **uses** 0 or more *Instrument* - A Data Collection Tool uses a Data Collection Mode *(relationship)*

Analysis Tool

- *Analysis Tool* **extends** *Tool* - An Analysis Tool extends Tool
- Zero or more *Analysis Tool* **applies** 1 or more *Instrument* - An Analysis Tool Applies an Instrument *(relationship)*

Processing Tool

- *Processing Tool* **extends** *Tool* - An Processing Tool extends Tool
- Zero or more *Processing Tool* **applies** 1 or more *Instrument* - An Processing Tool Applies an Instrument *(relationship)*

Data Collection Method

- *Data Collection Method* **extends** *Method* - An Data Collection Method extends Method

Analysis Method

- *Analysis Method* **extends** *Method* - An Analysis Method extends Method

Processing Method

- *Processing Method* **extends** *Method* – A Processing Method extends Method

Data Collection Instrument

- *Data Collection Instrument* **extends** *Instrument* – A Data Collection Instrument extends Instrument
- Zero or more *Data Collection Instrument* **implements** 1 *Data Collection Method* – A Data Collection Instrument implements a Data Collection Method *(relationship)*

Analysis Instrument

- *Analysis Instrument* **extends** *Instrument* – An Analysis Instrument extends Instrument
- Zero or more *Analysis Instrument* **implements** 1 *Analysis Method* – An Analysis Instrument implements an Analysis Method *(relationship)*

Processing Instrument

- *Processing Instrument* **extends** *Instrument* – A Processing Instrument extends Instrument
- Zero or more *Processing Instrument* **implements** 1 *Processing Method* – A Processing Instrument implements a Processing Method *(relationship)*

Data Collection Mode

- Zero or more *Data Collection Mode* **enables** 0 or more *Data Collection Instrument* – A Data Collection Mode enables a Data Collection Instrument *(relationship)*

SegmentOfText

- *SegmentOfText* **extends** *Physical Segment* – SegmentOfText extends Physical Segment
- One *SegmentOfText* **has** 0 or more *LineParameter* – SegmentOfText has a LineParameter *(relationship)*
- One *SegmentOfText* **has** 0 or more *CharacterParameter* – SegmentOfText has a CharacterParameter describing the location of the segment *(relationship)*

LineParameter (not the source for any relationship)

CharacterParameter (not the source for any relationship)

SegmentOfXML

- *SegmentOfXML* **extends** *Physical Segment* – SegmentOfXML extends Physical Segment
- One *SegmentOfXML* **has** 0 or more *Node* – SegmentOfXML has a Node describing the location of the segment *(relationship)*

Node (not the source for any relationship)

SegmentOfAudio

- *SegmentOfAudio* **extends** *Physical Segment* – SegmentOfAudio extends Physical Segment
- One *SegmentOfAudio* **has** 0 or more *BeginEnd* – SegmentOfAudio has a BeginEnd describing the location of the segment *(relationship)*

BeginEnd (not the source for any relationship)

SegmentOfVideo

- *SegmentOfVideo* **extends** *Physical Segment* – SegmentOfVideo extends Physical Segment
- One *SegmentOfVideo* **has** 0 or more *BeginEnd* – SegmentOfVideo has a BeginEnd describing the location of the segment

SegmentOfImage

- *SegmentOfImage* **extends** *Physical Segment* – SegmentOfImage extends Physical Segment
- One *SegmentOfImage* **has** 0 or more *Area* – SegmentOfImage has an Area describing the location of the segment *(relationship)*

Area (not the source for any relationship)

SegmentOfTriplestore

- *SegmentOfTriplestore* **extends** *Physical Segment* – SegmentOfTriplestore extends Physical Segment

SegmentOfRelationalDatabase

- *SegmentOfRelationalDatabase* ***extends*** *Physical Segment* – SegmentOfRelationalDatabase extends Physical Segment
- *SegmentOfRelationalDatabase* **has** *SQLselection* – SegmentOfRelationalDatabase is described by an SQL query *(relationship)*

SQLselection (not the source for any relationship)

SegmentOfWebPage

- *SegmentOfWebPage* **extends** *Physical Segment* – SegmentOfWebPage extends Physical Segment
- One *SegmentOfWebPage* **has** 0 or more *Node* – SegmentOfWebPage has a Node describing the location of the segment (like SegmentOfXML) *(relationship)*

SegmentOfDDIDescribedData

- *SegmentOfDDIDescribedData* **extends** *Physical Segment* – SegmentOfDDIDescribedData extends Physical Segment

SegmentOfMultidimensional

- *SegmentOfMultidimensional* **extends** *Physical Segment* – SegmentOfMultidimensional extends Physical Segment
- One *SegmentOfMultidimensional* **has** 1 or more *BoundingCube* – SegmentOfMultidimensional has one or more (hyper)BoundingCubes describing the values of the two tuples defining the location of the segment of the data object

BoundingCube

- One *BoundingCube* **has 0** or more *DimensionNames* – BoundingCube has zero or more DimensionNames which label the levels of a single dimension. An example might be the "book" dimension might have labels "Genesis, Exodus"….

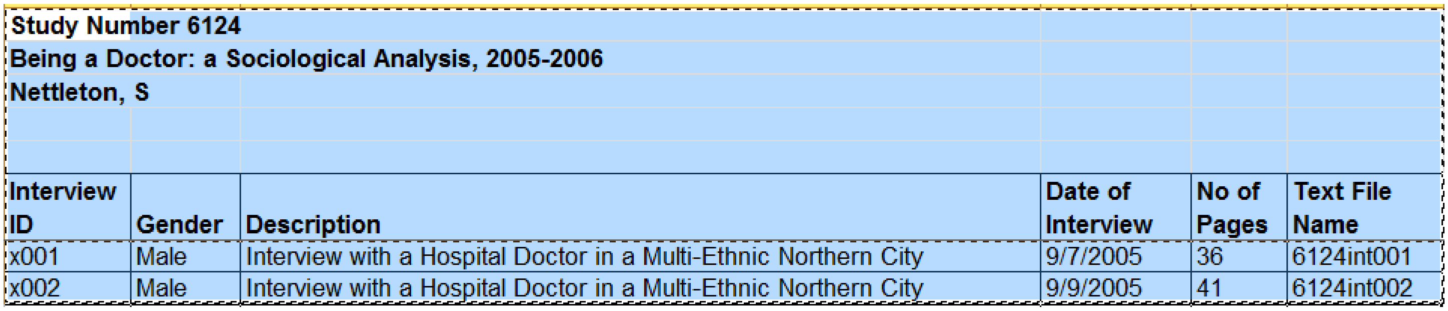
DimensionNames (not the source for any relationship)

Walk through of Selected Use Cases

UKDA Being a Doctor

The Bergen group looked at UKDA study number 6124 "Being a Doctor: a Sociological Analysis, 2005-2006" as an example of a qualitative data collection needing metadata at the object level only. This collection has extensive metadata at the study level: Principal Investigator, Sponsor, Distributed By, Bibliographic Citation, Acknowledgements, Disclaimer, Copyright and Access Control Information, Abstract, Topical Coverage, Spatial Coverage, Temporal Coverage, Universe Sampled, Methodology Description, and Language.

Metadata also include a table (as an Excel spreadsheet), part of which is shown below, listing each file in the collection along with some attributes for the file.



Study Number 6124					
Being a Doctor: a Sociological Analysis, 2005-2006					
Nettleton, S					
Interview ID	**Gender**	**Description**	**Date of Interview**	**No of Pages**	**Text File Name**
x001	Male	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	9/7/2005	36	6124int001
x002	Male	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	9/9/2005	41	6124int002

In our model each row of the spreadsheet corresponds to a Logical Resource, with the text file listed in the last column being a Physical Storage Instance. If an audio file of the same interview were listed on the same line - that would be a separate Physical Storage Instance linked to the same Logical Resource.

In the model a Logical Resource has these properties: Name, Label, Description, Type, IsOriginal, Digital Version Available, Creation Date, Citation, Temporal Coverage, Spatial Coverage, Topical Coverage, Identifiable, Versionable

In this example the entry in the *Description* column would be entered into the Description attribute. *The Date of Interview* column could be entered into the Temporal Coverage attribute if the interview was focused on the current time period at the time of the interview. If the interview was retrospective, focused on some other time period, then that time period would be the Temporal Coverage.

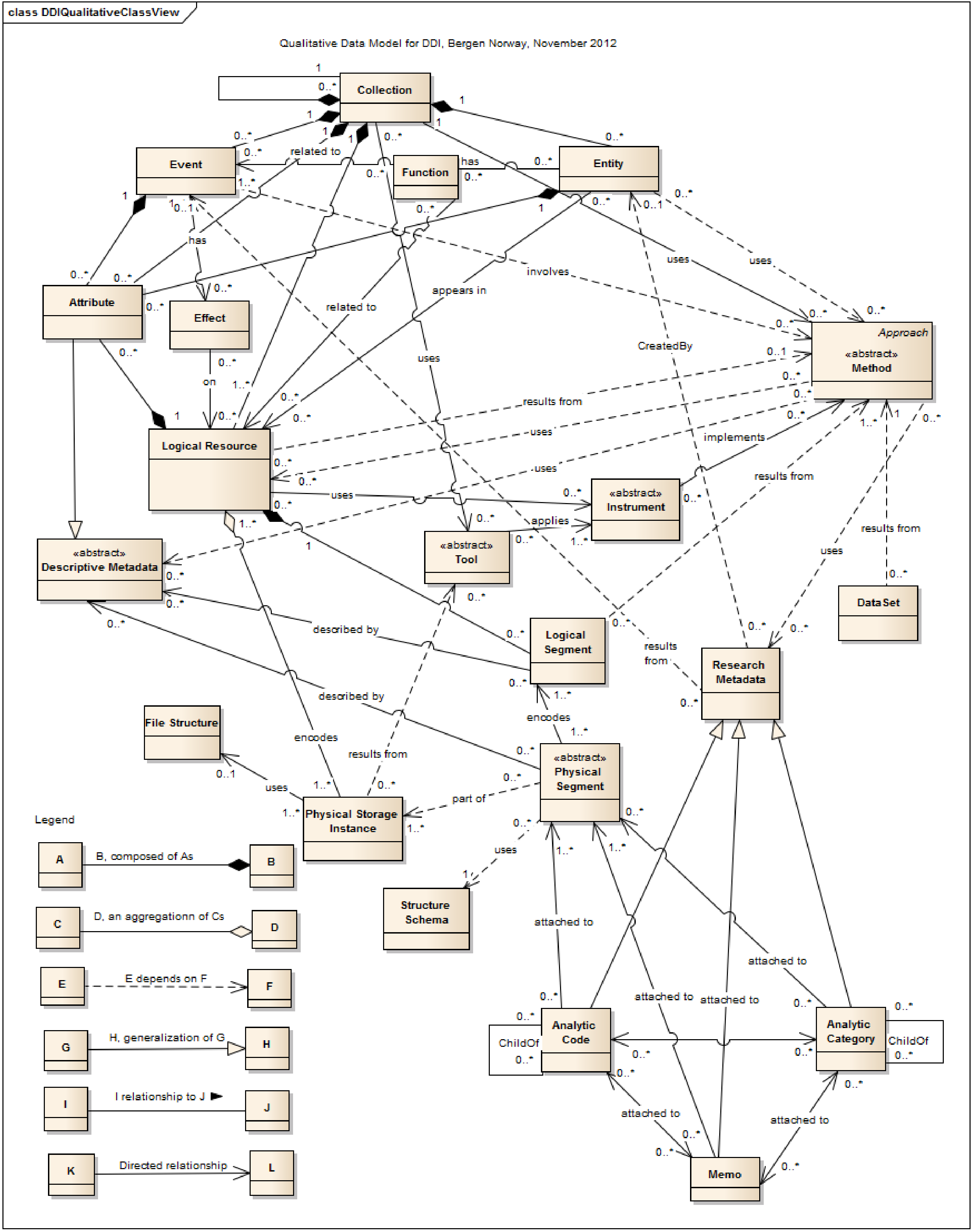
As an Identifiable DDI element, there would be an inherited UserID attribute which would be the appropriate element in which to store the *Interview ID* value.

While *Gender* and *No. Of Pages* could certainly be included in a description, these would not be explicitly searchable, nor would it be machine-actionable to recreate the spreadsheet grid from DDI metadata. This is where the Attribute element in the model comes into play. Two Attribute elements could be attached to each Logical Resource, one with a Name of Gender and one with a name of Pages (and a Label of No of Pages). These would correspond to the appropriate columns in the spreadsheet and would have the appropriate entries in the corresponding Value elements. So for the first Row above:

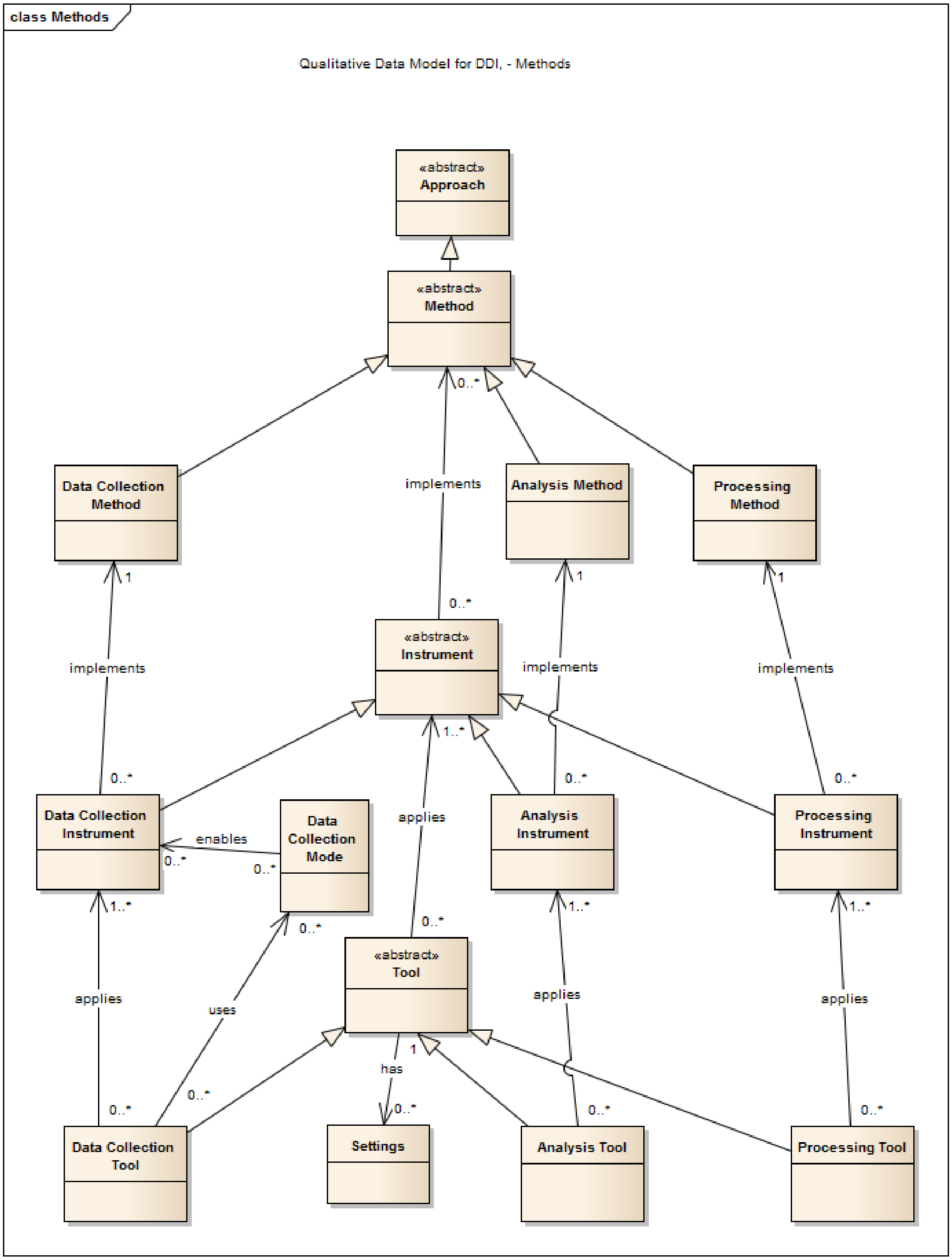
A **Logical Resource** with a

> *UserID* of "x001"
>
> *Description* of "Interview with a Hospital Doctor in a Multi-Ethnic Northern City"
>
> Temporal Coverage of 9/7/2005 or the time period about which the interview was concerned (see above)
>
> An **Event** with properties of
>
> > **Type** of "Interview" and an
> >
> > **Event Date Time** of 9/7/2005
>
> An **Entity** with both
>
> > an **Attribute** having
> >
> > > *Name* of Gender
> > >
> > > *Value* of Male
> >
> > a **Function** having a type of "Interviewee"
>
> A **SegmentOfText** with an **Attribute** having
>
> > *Name* of Pages
> >
> > *Label* of No of Pages
> >
> > *Value* of 36
> >
> > (Note that this Segment of Text would be the whole interview. References to lines and pages of text are a complicated issue with digital text objects in certain formats like word processing program file formats. Changes in the version of the program or with different default paper sizes opening the file might cause different folding of the lines or pagination. This is an area of attention for careful curation.)
>
> A **Physical Instance** describing the file 6124int001 (location, layout etc.)

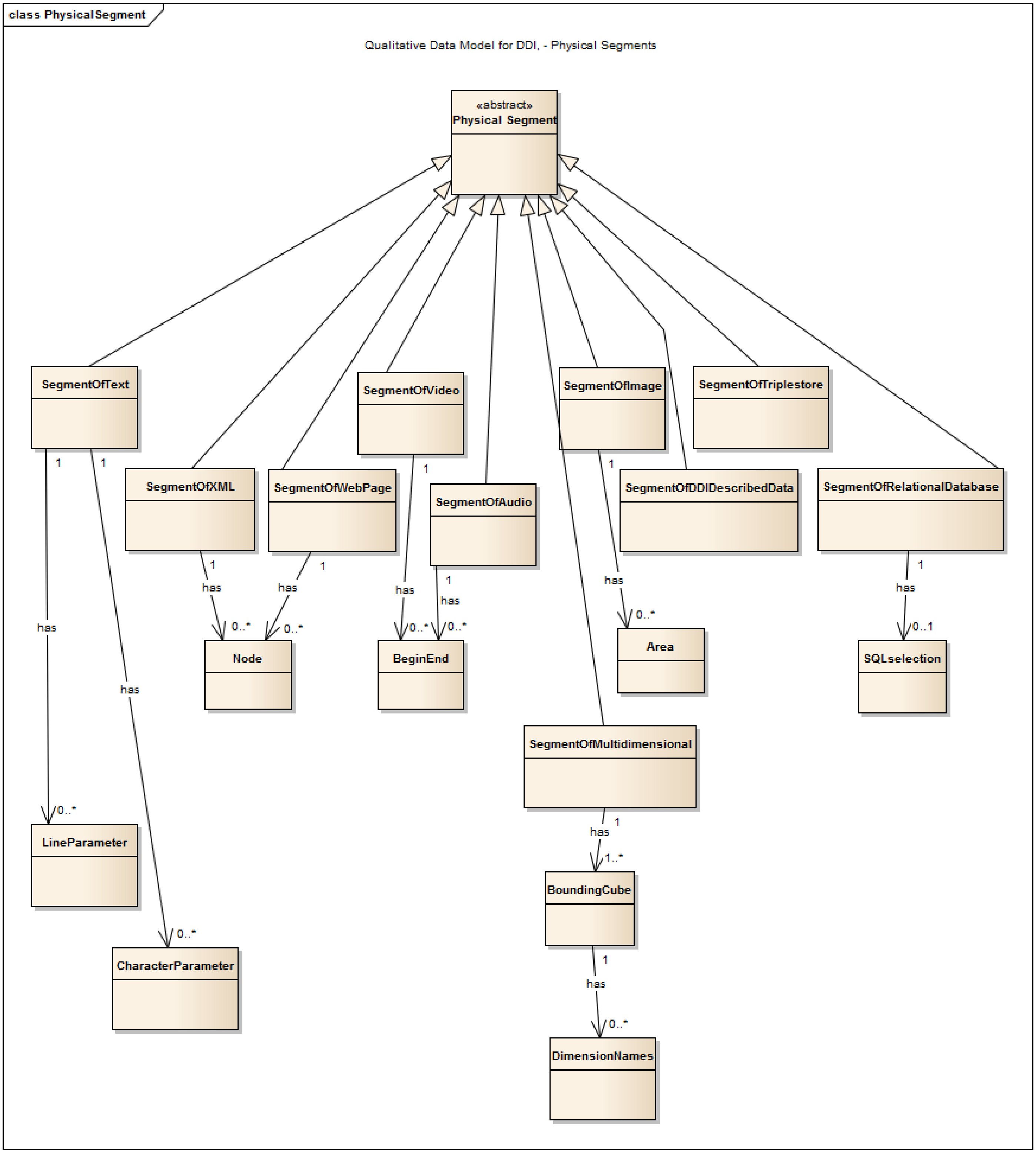
Model Element Relationships (Diagrams)



Methods


class Methods
Qualitative Data Model for DDI, - Methods
«abstract»
Approach
«abstract»
Method
0..*
Data Collection Method
implements
Analysis Method
Processing Method
1
1
1
0..*
«abstract»
Instrument
implements
implements
implements
1..*
0..*
0..*
0..*
Data Collection Instrument
enables
Data Collection Mode
applies
Analysis Instrument
Processing Instrument
0..*
0..*
0..*
0..*
1..*
1..*
1..*
«abstract»
Tool
applies
uses
applies
applies
1
has
0..*
0..*
0..*
0..*
0..*
Data Collection Tool
Settings
Analysis Tool
Processing Tool

Physical Segments



Challenges

Many Ways to Describe a Segment

The model described here includes several methods for describing the location of segments within a variety of digital objects. All of these methods rely on the object having a certain homogeneity of content. Each method locates only one type of material – text, or video, or XML. Complex digital objects may contain multiple types of material, or embed one type of material in a complex structure. A word processing file may contain text and images, or have text in a body and frames within that body. A Web page may contain text and an embedded video segment and an interactive program. The DDI Qualitative Data Model may need to address the more common types of more complex objects. The model may also need to offer some more extensible method of describing segments.

Even for the types of objects the model describes here, there can be difficulties which arise as the objects are displayed. For simple word processing document files, for example, values of line number and character within line for a LineParameter segment description change as the paper size setting changes. This method depends on the document being stored in a format that does not allow changes to layout. Documentation of the original document settings (like page size) which determine the line endings is also important.

Remaining Issues

One large remaining task is to integrate this model with the existing (and future) DDI Lifecycle models. Much of the metadata at the StudyUnit level will be relevant (e.g., Universe, Sampling, Embargo, Purpose, etc.). These metadata may be appropriate at multiple levels in the Qualitative model (e.g., Collection, Logical Resource, and perhaps Logical Segment), necessitating additional properties and perhaps even new objects.

Questions

Collection, Entity, Event, Logical Resource can have an Attribute (having a name value pair). Should this take advantage of the rest of DDI and allow for use by reference of category and code schemes to allow for comparability across studies? Isn't this just a value for a variable?

The modeling group discussed whether the model should include properties of a Physical Storage Instance like total file size in bytes, or pixel dimensions for an image. Whether to include these properties in the DDI model or to leave them to other standards (like, PREMIS) has not been decided.

Are our relationship definitions redundant? Do we need better definitions focusing on the verb rather than just repeating the triple?

We should review our choice of modeling elements and verbs in the relationships

- e.g., "Logical Resource is composed of Attributes" is modeled as a composition (is composed of?)
- e.g., "Segment of XML has Node" is modeled as a relationship

Appendix A – Participants

Initial Working Group

Eirik Alvær	Norwegian Social Science Data Service, Bergen
Atle Alvheim	Norwegian Social Science Data Service, Bergen
Iris Braun	Archive, Bielefeld University
Angad Bhat	UK Data Archive, Essex
Louise Corti	UK Data Archive, Essex
Anne Sofie Fink	Danish Data Archive, Denmark
Damien Gallagher	Irish National Archives, Dublin
Tobias Gebel	Archive, Bielefeld University
Arofan Gregory	Metadata Technology, USA
Russel Hathaway	Inter-university Consortium for Political and Social Research, Michigan
Jani Hautamaki	Finnish Social Science Data Service, Tampere
Wolfram Horstmann	University of Beilefeld
Larry Hoyle	Institute for Policy and Social Research, University of Kansas
Rene Van Horik	Data Archiving and Networked Services, Hague
Herve l_Hours	UK Data Archive, Essex
Arja Kuula	Finnish Social Science Data Service, Tampere
Jeremy Iverson	Colectica, US
Arja Kuula	Finnish Social Science Data Service, Tampere
Kaye Marz	Inter-university Consortium for Political and Social Research, Michigan
Agustina Martinez	Liverpool John Moores University
Steve McEachern	Australian Data Archive, Australia
Barry Radler	Institute on Aging, Wisconsin-Madison
Ben Ryan	Timescapes, Leeds
Andrea Smioski	WISDOM, Austria
Dan Smith	Colectica, US
Joachim Wackerow	GESIS-Leibniz Institute for the Social Sciences, Mannheim
Cornelia Zuell	GESIS-Leibniz Institute for the Social Sciences, Germany

Working Group Meeting, Gothenburg, Sweden, 2011

Sofia Arvidsson	Swedish Social Science Data Service, Sweden
Louise Corti	UK Data Archive, Essex (Co-Organizer)
Anne Sofie Fink	Danish Data Archive, Denmark
Damien Gallagher	Irish National Archives, Dublin
Guillaume Garcia	Quali BE, Research Quetelet
Tobias Gebel	Archive, Bielefeld University
Arofan Gregory	Metadata Technology, USA (Co-Organizer)
Larry Hoyle	Institute for Policy and Social Research, University of Kansas
Jani Hautamaki	Finnish Social Science Data Service, Tampere
Agustina Martinez	Liverpool John Moores University

Steven McEachern	Australian Data Archive, Australia
Rene Van Horik	Data Archiving and Networked Services, Hague
Eirik Vestrheim	Norwegian Social Science Data Service, Bergen
Christian Meier zu Verl	Archive, Bielefeld University
Johanna Vompras	Archive, Bielefeld University
Joachim Wackerow	GESIS-Leibniz Institute for the Social Sciences, Mannheim (Co-Organizer)
Cornelia Zuell	GESIS, Germany

Modeling Qualitative Metadata Structures and the DDI, 28-30 November 2012, Bergen, Norway

Eirik Alvær	Norwegian Social Science Data Service, Bergen
Noemi Betancort	Qualiservice, Bremen
Iris Braun	Archive, Bielefeld University
Louise Corti	UK Data Archive, Essex (Co-Organizer)
Damien Gallagher	Irish National Archives, Dublin
Tobias Gebel	Archive, Bielefeld University
Arofan Gregory	Metadata Technology, USA (Co-Organizer)
Jani Hautamaki	Finnish Social Science Data Service
Larry Hoyle	Institute for Policy and Social Research, University of Kansas
Arja Kuula	Finnish Social Science Data Service
Agustina Martinez	Liverpool John Moores University
Joachim Wackerow	GESIS-Leibniz Institute for the Social Sciences, Mannheim (Co-Organizer)
Cornelia Zuell	GESIS-Leibniz Institute for the Social Sciences, Mannheim

Others

Mary Vardigan provided extensive copyediting on this document.

Thanks to Anna Tatarko for slogging through listing all of the relationships in the model.