Overview of DDI

Arofan Gregory

ICH

November 18, 2011

Credits

- The slides were developed for several DDI workshops at IASSIST conferences and at GESIS training in Dagstuhl/Germany
- Major contributors
  - Wendy Thomas, Minnesota Population Center
  - Arofan Gregory, Open Data Foundation
- Further contributors
  - Joachim Wackerow, GESIS – Leibniz Institute for the Social Sciences
  - Pascal Heus, Open Data Foundation

Overview

- Background and Introduction
- DDI Content – High Level
- DDI Typical Use Cases
- DDI Structural Components
- Additional Technical Topics
- Overview of On-Going Activities Related to DDI, SDMX, and GSBPM

The Data Documentation Initiative

- The Data Documentation Initiative is an XML specification to capture structured metadata about “microdata” (broad sense)
- First generation DDI 1.0…2.1 (2000-2008)
  - Focus on single archived instance
- Second generation DDI 3.0 (2008)
  - Focus on life cycle
  - Go beyond the single survey concept
- Governance: DDI Alliance
  - Membership based organizations (35 members)
  - Data archives, producers, research data centers, university data libraries, statistics organizations
  - http://www.ddialliance.org/org/index.html

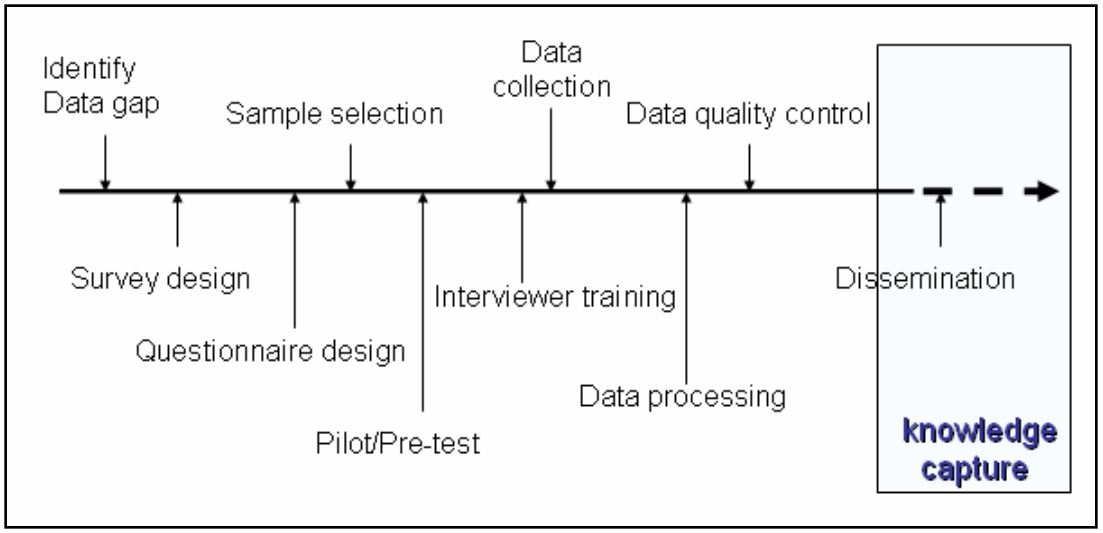
DDI Timeline / Status

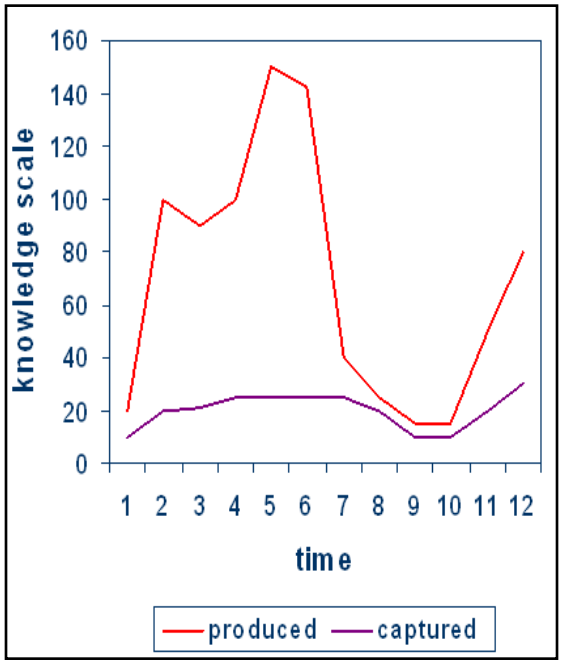
- Pre-DDI 1.0
  - 70's / 80's OSIRIS Codebook
  - 1993: IASSIST Codebook Action Group
  - 1996 SGML DTD
  - 1997 DDI XML
  - 1999 Draft DDI DTD
- 2000 – DDI 1.0
  - Simple survey
  - Archival data formats
  - Microdata only
- 2003 – DDI 2.0
  - Aggregate data (based on matrix structure)
  - Added geographic material to aid geographic search systems and GIS users
- 2003 – Establishment of DDI Alliance
- 2004 – Acceptance of a new DDI paradigm
  - Lifecycle model
  - Shift from the codebook centric / variable centric model to capturing the lifecycle of data
  - Agreement on expanded areas of coverage
- 2005
  - Presentation of schema structure
  - Focus on points of metadata creation and reuse
- 2006
  - Presentation of first complete 3.0 model
  - Internal and public review
- 2007
  - Vote to move to Candidate Version (CR)
  - Establishment of a set of use cases to test application and implementation
  - October 3.0 CR2
- 2008
  - February 3.0 CR3
  - March 3.0 CR3 update
  - April 3.0 CR3 final
  - April 28th 3.0 Approved by DDI Alliance
  - May 21st DDI 3.0 Officially announced
  - Initial presentations at IASSIST 2008
- 2009
  - DDI 3.1 approved in May
  - Ongoing work on sampling and survey design, documenting data quality, qualitative data, and other features

DDI 3.0

The life cycle

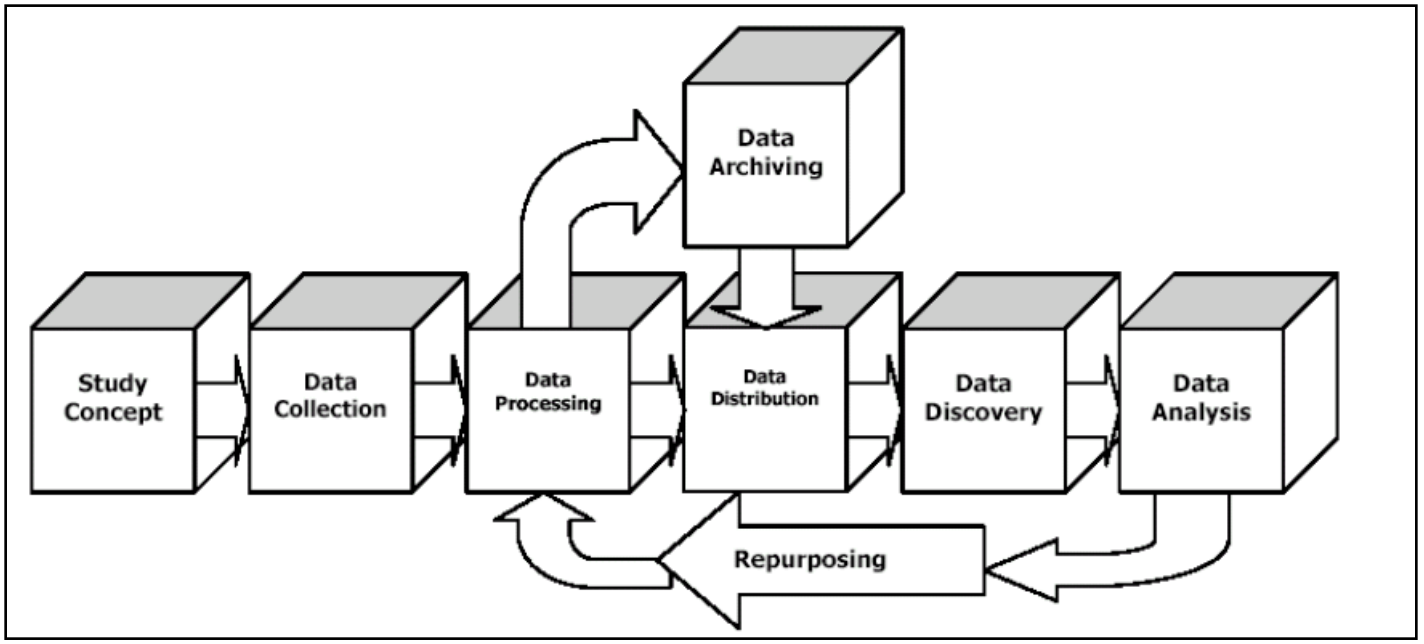
When to capture metadata?





- Metadata must be captured at the time the event occurs!
- Documenting after the facts leads to considerable loss of information
- Multiple contributors are typically involved in this process (not only the archivist)
- Metadata should be used to automate *throughout* the entire process
- This is true for producers and researchers

DDI 3.0 and the Survey Life Cycle



- A survey is not a static process: It dynamically evolved across time and involves many agencies/individuals
- DDI 2.x is about archiving, DDI 3.0 across the entire “life cycle”
- 3.0 focus on metadata reuse (minimizes redundancies/discrepancies, support comparison)
- Also supports multilingual, grouping, geography, and others
- 3.0 is extensible

Requirements for 3.0

- Improve and expand the **machine-actionable** aspects of the DDI to support programming and software systems
- **Support CAI instruments** through expanded description of the questionnaire (content and question flow)
- Support the **description of data series** (longitudinal surveys, panel studies, recurring waves, etc.)
- Support **comparison**, in particular comparison by design but also comparison-after-the fact (harmonization)
- Improve support for describing **complex data files** (record and file linkages)
- Provide **improved support for geographic content** to facilitate linking to geographic files (shape files, boundary files, etc.)

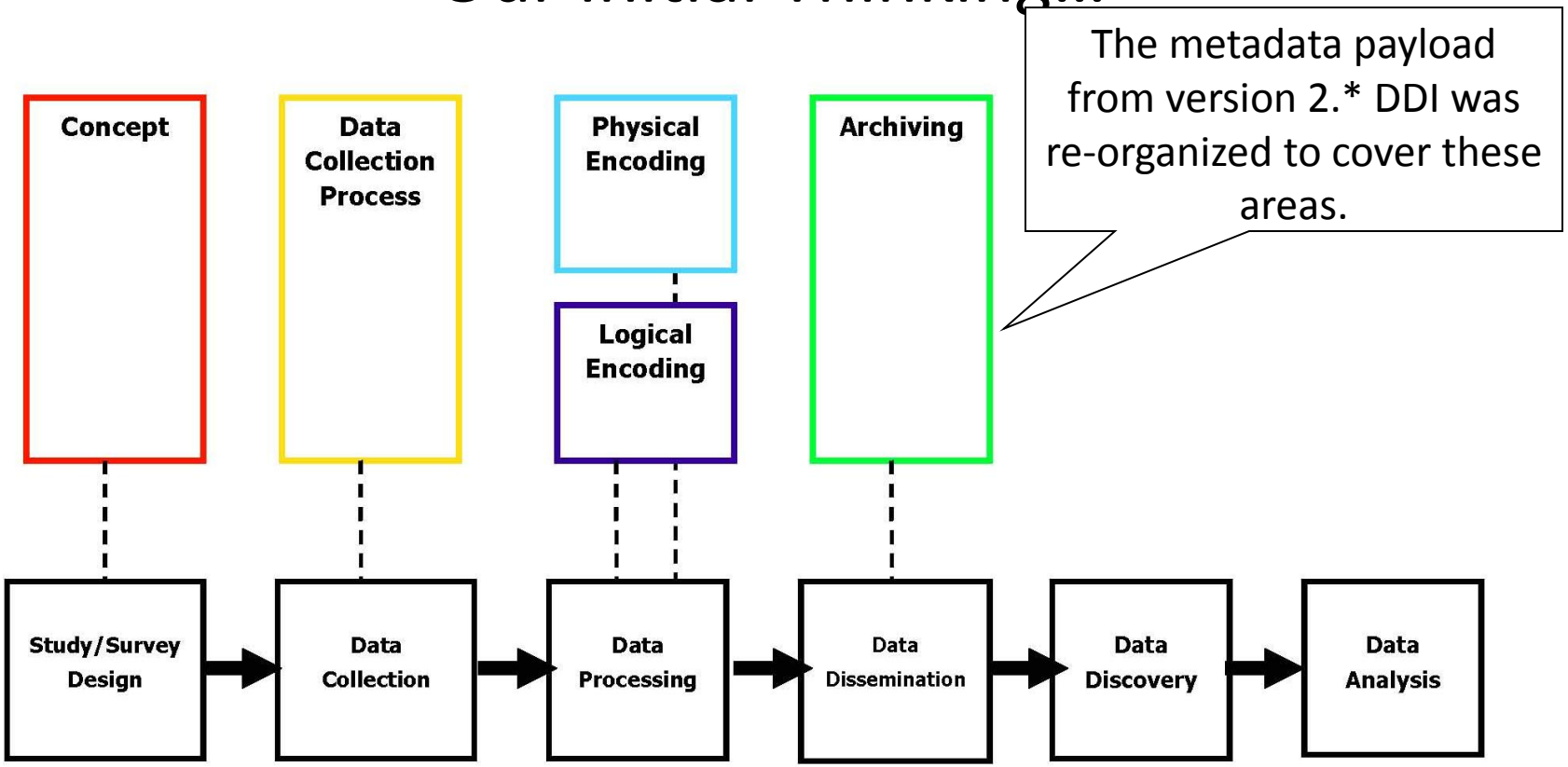
Approach

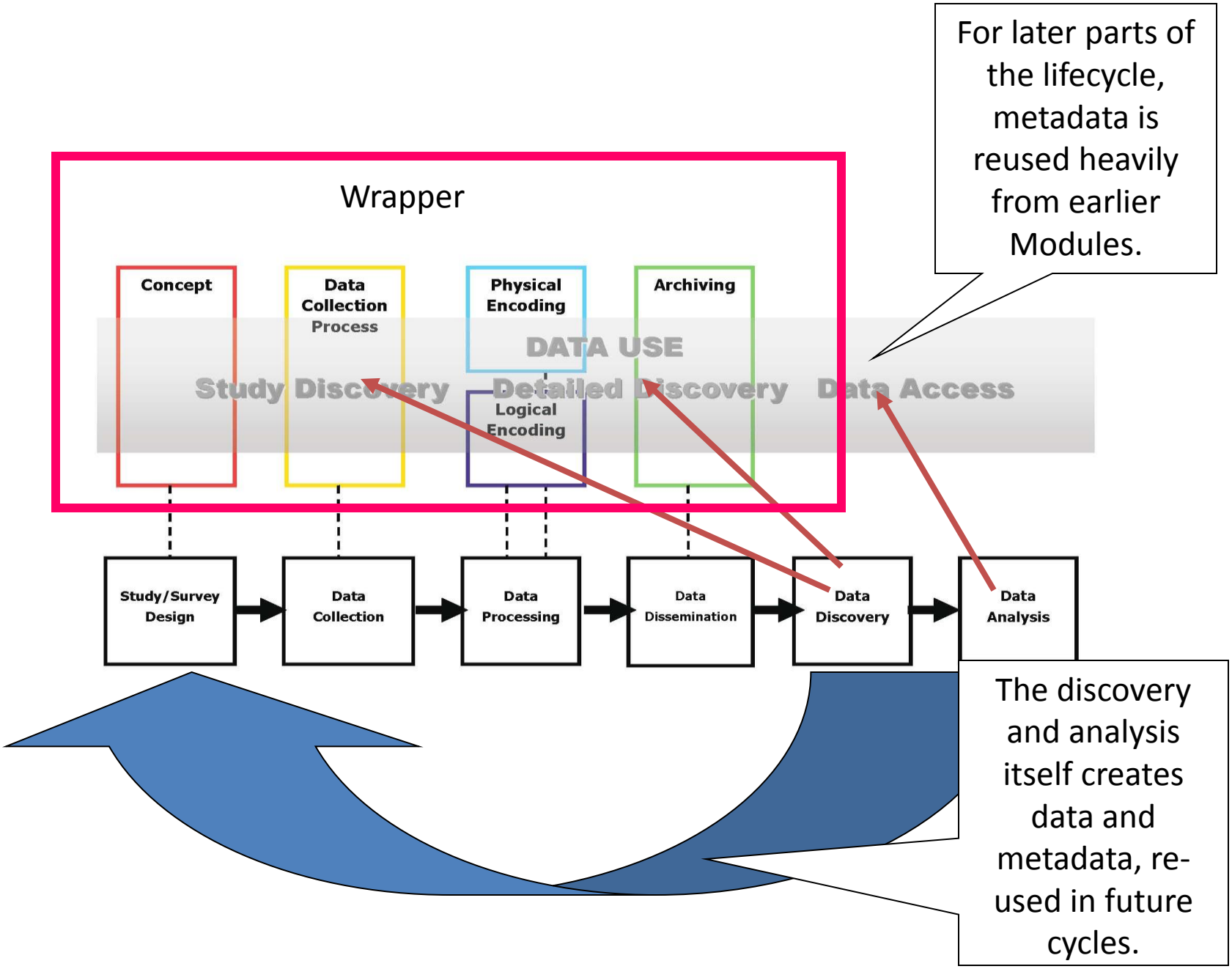
- **Shift from** the **codebook** centric model of early versions of DDI **to a lifecycle model**, providing metadata support from data study conception through analysis and repurposing of data
- **Shift** from an XML Data Type Definition (DTD) to an **XML Schema** model to support the lifecycle model, ***reuse of content*** and increased controls to support programming needs
- Redefine a "single **DDI instance**" to include a "***simple instance***" similar to DDI 1/2 which covered a single study and "***complex instances***" covering ***groups*** of related studies. Allow a single study description to contain **multiple data products** (for example, a microdata file and aggregate products created from the same data collection).
- Incorporate the requested functionality in the first published edition

Designing to support registries

- Resource package
  - structure to publish non-study-specific materials for reuse (concepts, classifications, questions,…)
- Extracting specified types of information into maintainable schemes
  - Universe, Concept, Category, Code, Question, Instrument, Variable, etc.
  - Very much like relational database tables
- Allowing for either internal or external references
  - Can include other schemes by reference and select only desired items
- Providing Comparison Mapping
  - Target can be external harmonized structure

Our Initial Thinking…




For later parts of the lifecycle, metadata is reused heavily from earlier Modules.
Wrapper
Concept
Data Collection Process
Physical Encoding
Logical Encoding
Archiving
DATA USE
Study Discovery
Detailed Discovery
Data Access
Study/Survey Design
Data Collection
Data Processing
Data Dissemination
Data Discovery
Data Analysis
The discovery and analysis itself creates data and metadata, re-used in future cycles.

DDI Content

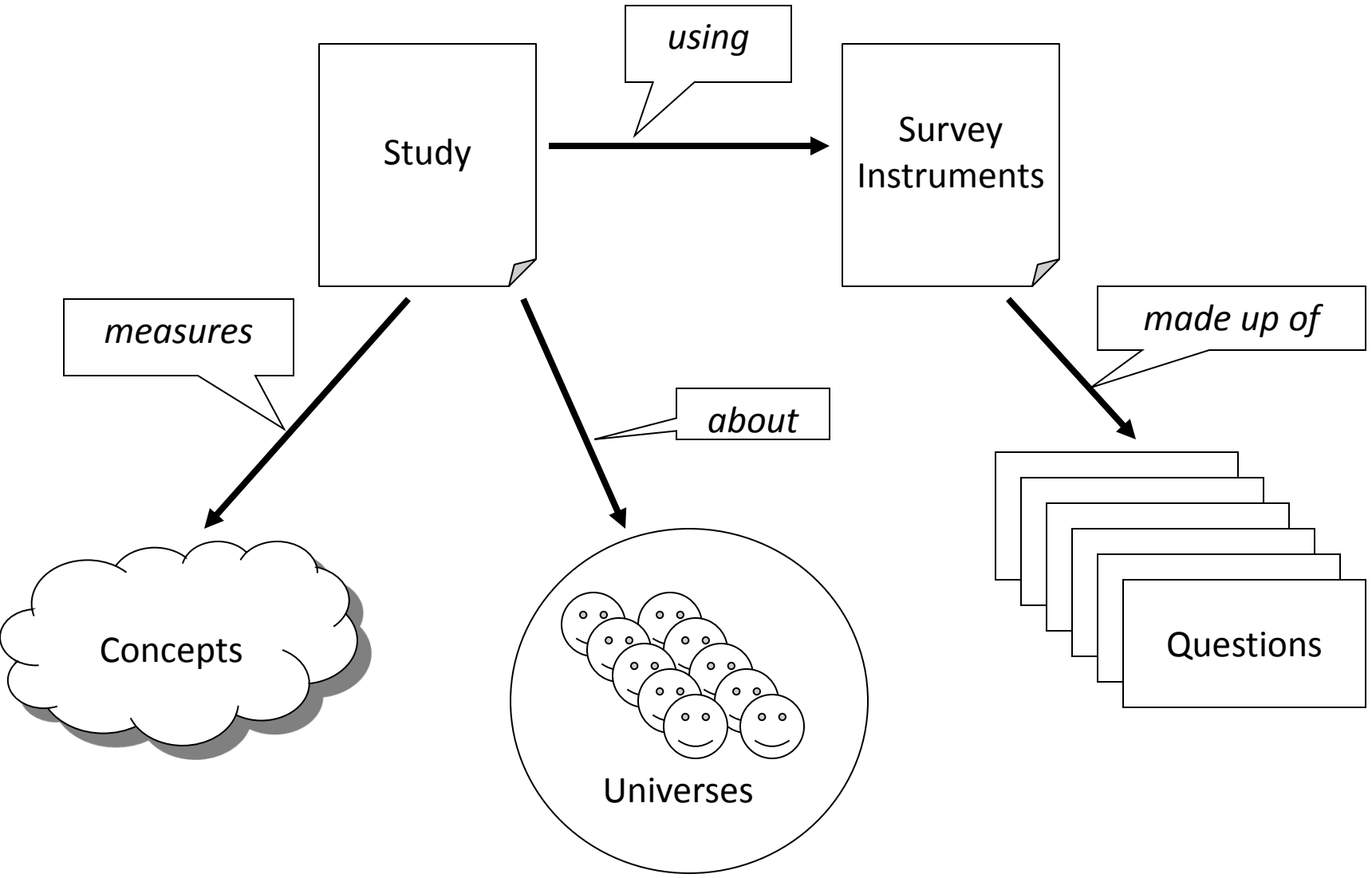
- DDI 3 may seem very technical
  - It is not an invention!
  - It is based on the metadata used across many different organizations for collecting, managing, and disseminating data
- This section introduces the types of metadata which are the content of DDI
  - Not a technical view, but a business view
  - You work with this metadata every day – it should be familiar to you
  - You may use different terminology

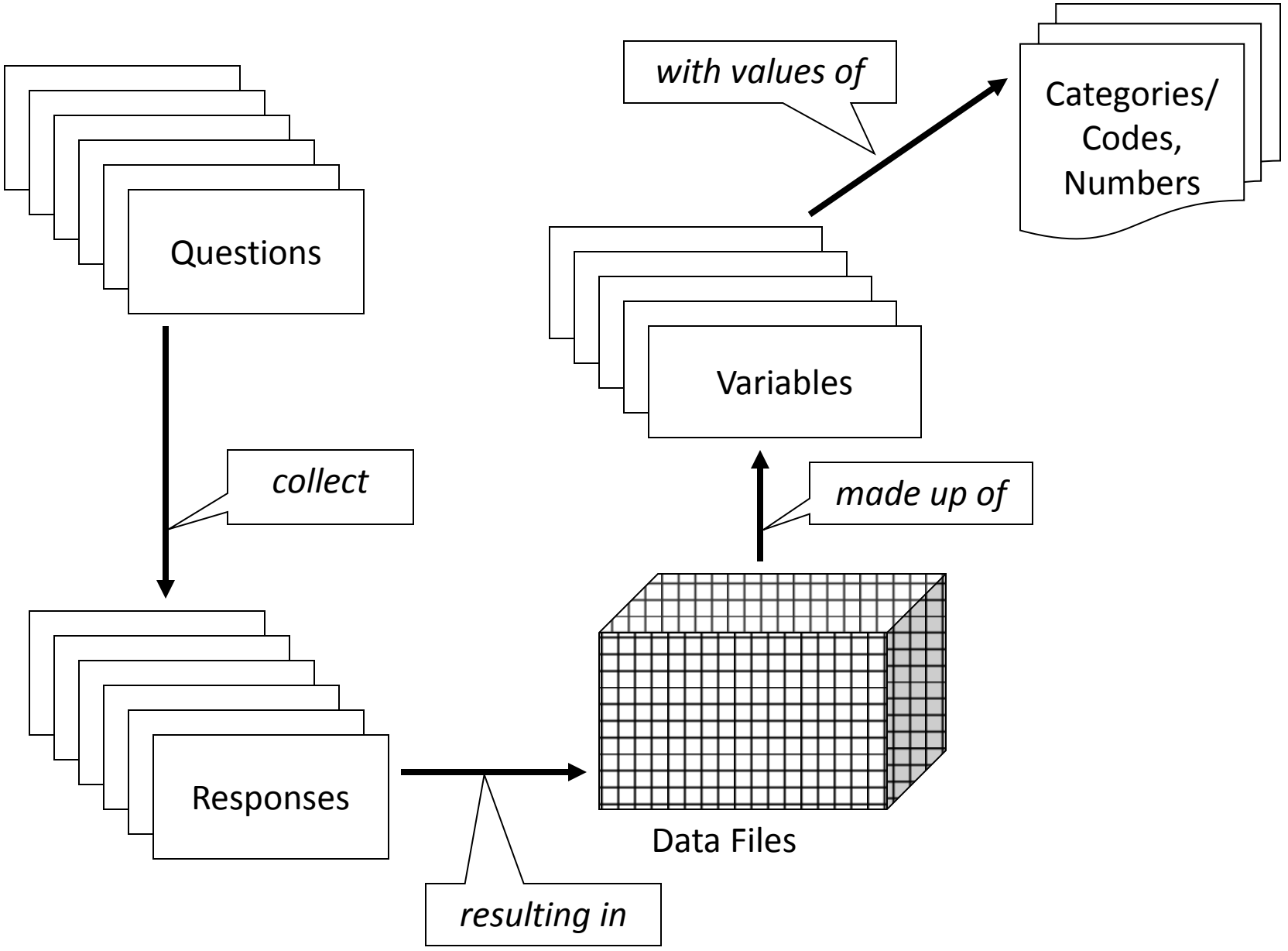
Basic Types of Metadata

- Concepts (“terms”)
- Studies (“surveys”, “collections”, “data sets”, “samples”, “censuses”, “trials”, “experiments”, etc.)
- Survey instruments (“questionnaire”, “form”)
- Questions (“observations”)
- Responses

Basic Types of Metadata (2)

- Variables (“data elements”, “columns”)
- Codes & categories (“classifications”, “codelists”)
- Universes (“populations”, “samples”)
- Data files (“data sets”, “databases”)


Study
using
Survey Instruments
measures
about
made up of
Concepts
Universes
Questions


Questions
collect
Responses
resulting in
Data Files
made up of
Variables
with values of
Categories/
Codes,
Numbers

Reuse Across the Lifecycle

- This basic metadata is reused across the lifecycle
  - Responses may use the same categories and codes which the variables use
  - Multiple waves of a study may re-use concepts, questions, responses, variables, categories, codes, survey instruments, etc. from earlier waves

Reuse by Reference

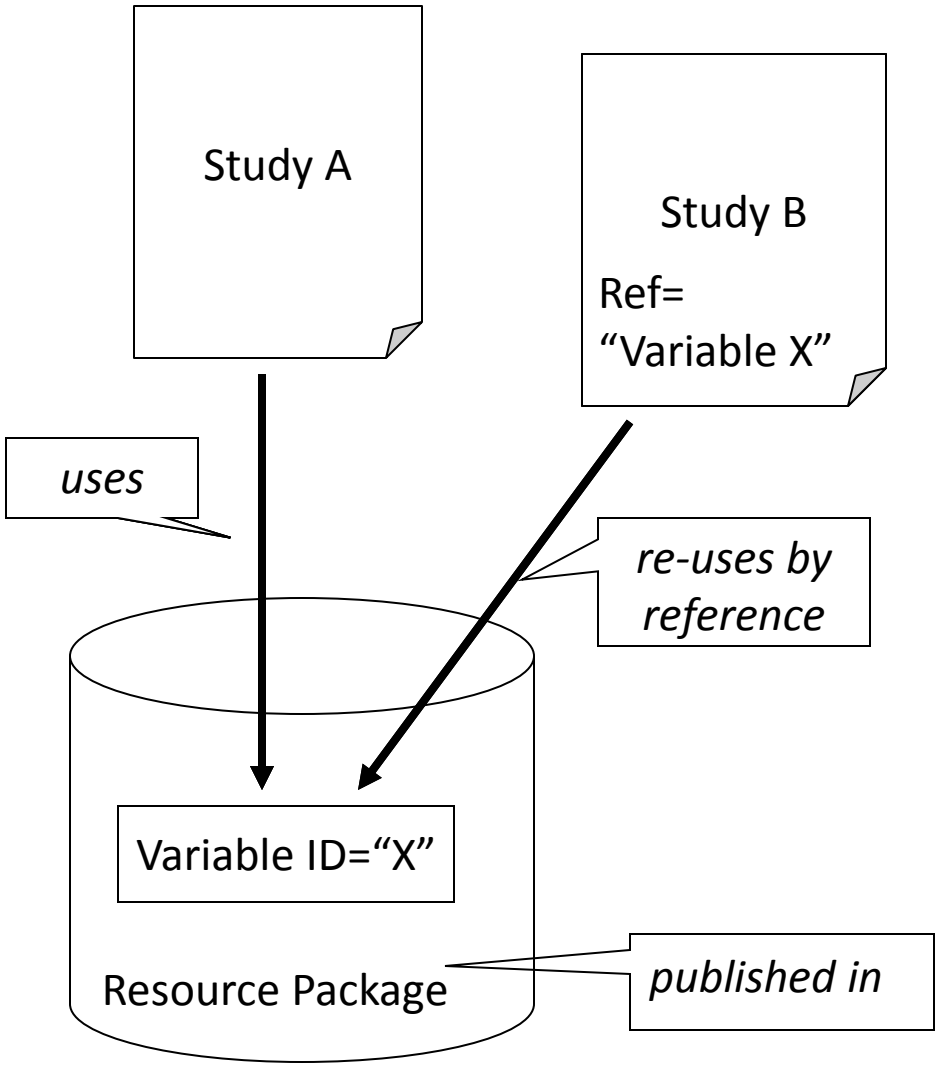
- When a piece of metadata is re-used, a *reference* can be made to the original
- In order to reference the original, you must be able to *identify* it
- You also must be able to *publish* it, so it is visible (and can be referenced)
  - It is published to the user community – those users who are allowed access

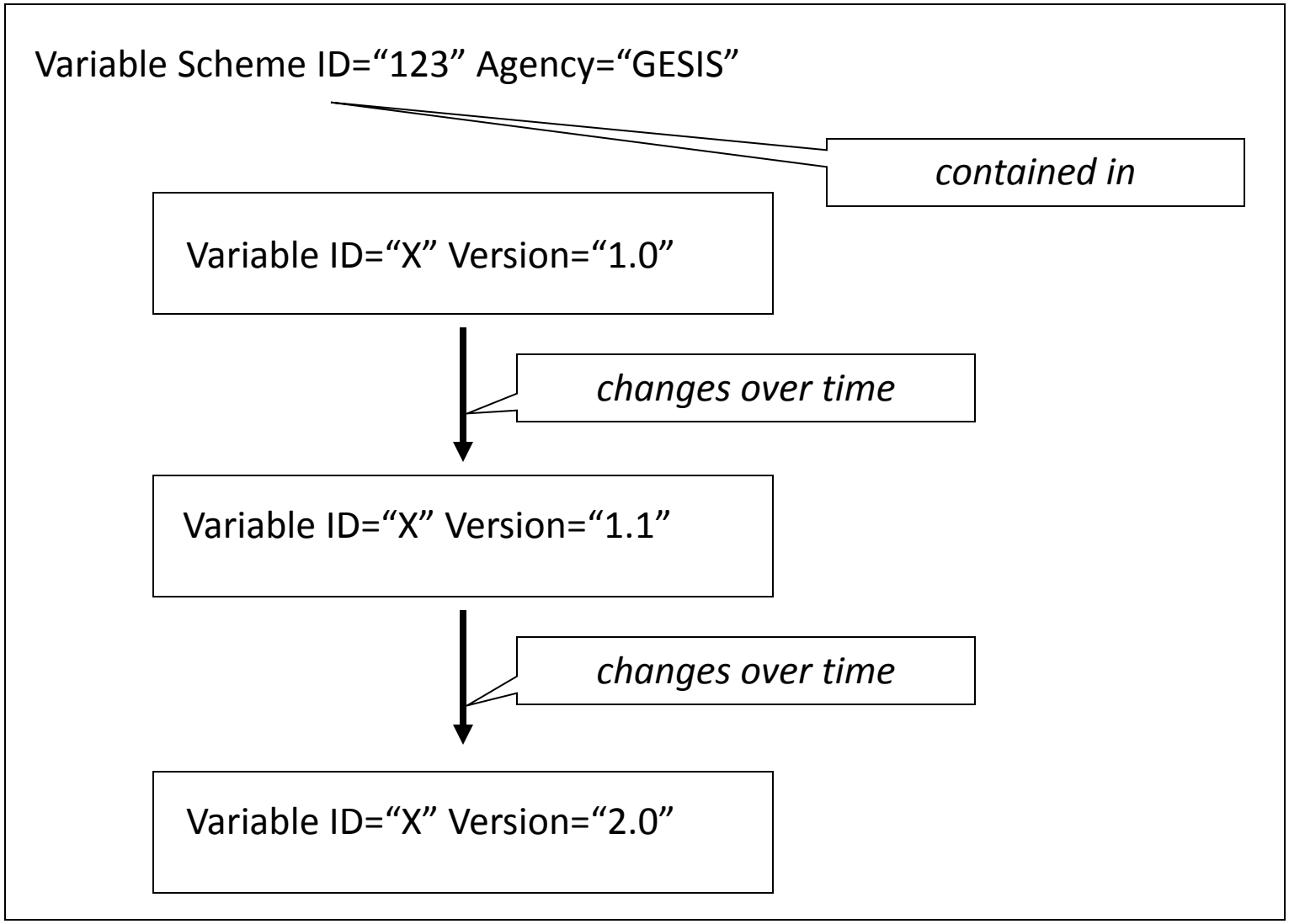
Change over Time

- Metadata items change over time, as they move through the data lifecycle
  - This is especially true of longitudinal/repeat cross-sectional studies
- This produces different *versions* of the metadata
- The metadata versions have to be *maintained* as they change over time
  - If you reference an item, it should not change: you reference a specific version of the metadata item

DDI Support for Metadata Reuse

- DDI allows for metadata items to be *identifiable*
  - They have unique IDs
  - They can be re-used by *referencing* those IDs
- DDI allows for metadata items to be *published*
  - The items are published in *resource packages*
- Metadata items are *maintainable*
  - They live in “schemes” (lists of items of a single type) or in “modules” (metadata for a specific purpose or stage of the lifecycle)
  - All maintainable metadata has a known owner or *agency*
- Maintainable metadata can be *versionable*
  - This reflects changes over time
  - The versionable metadata has a version number


Study A
Study B
Ref=
“Variable X”
uses
re-uses by reference
Variable ID=“X”
Resource Package
published in

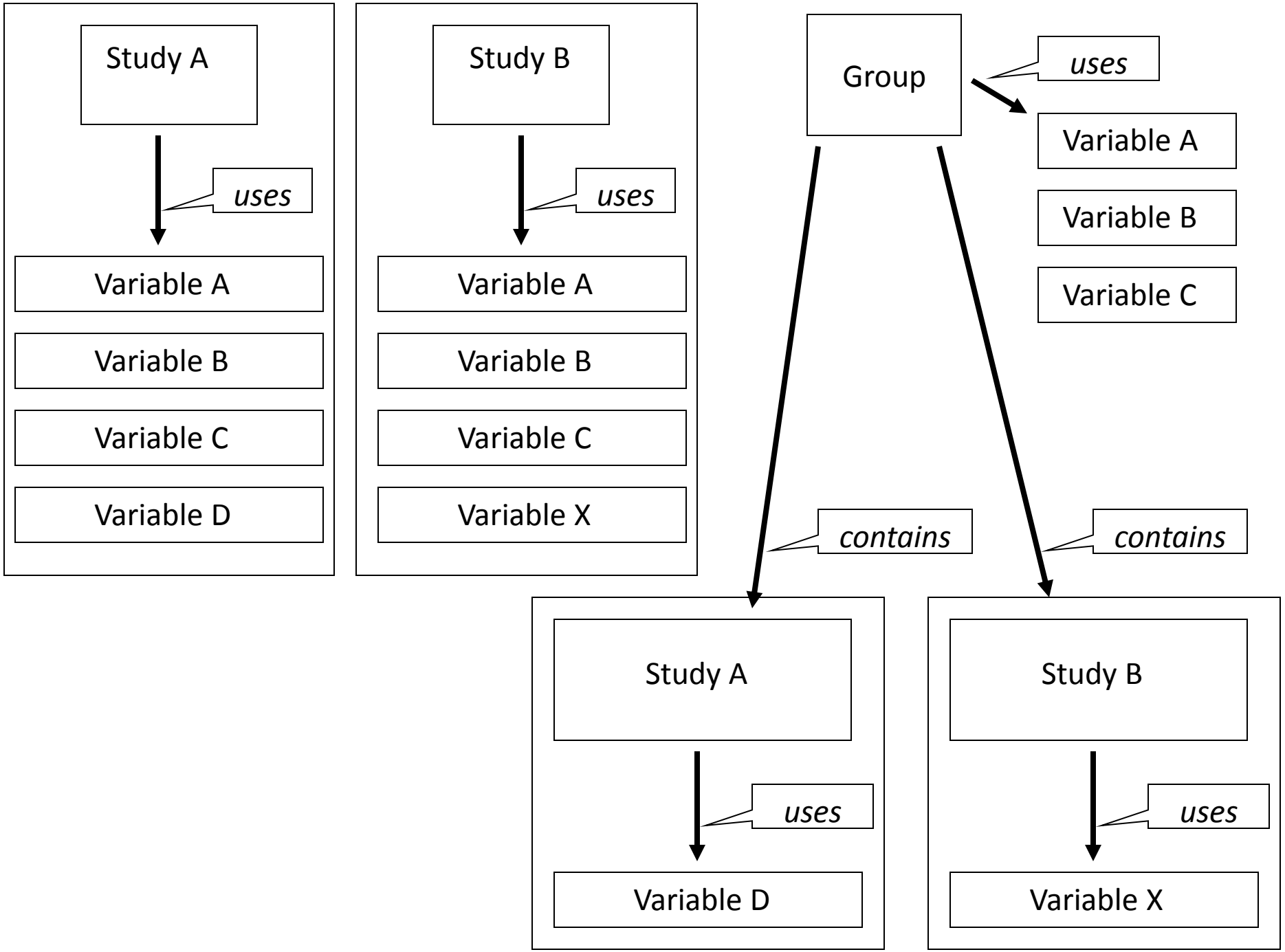

Variable Scheme ID="123" Agency="GESIS"
contained in
Variable ID="X" Version="1.0"
changes over time
Variable ID="X" Version="1.1"
changes over time
Variable ID="X" Version="2.0"

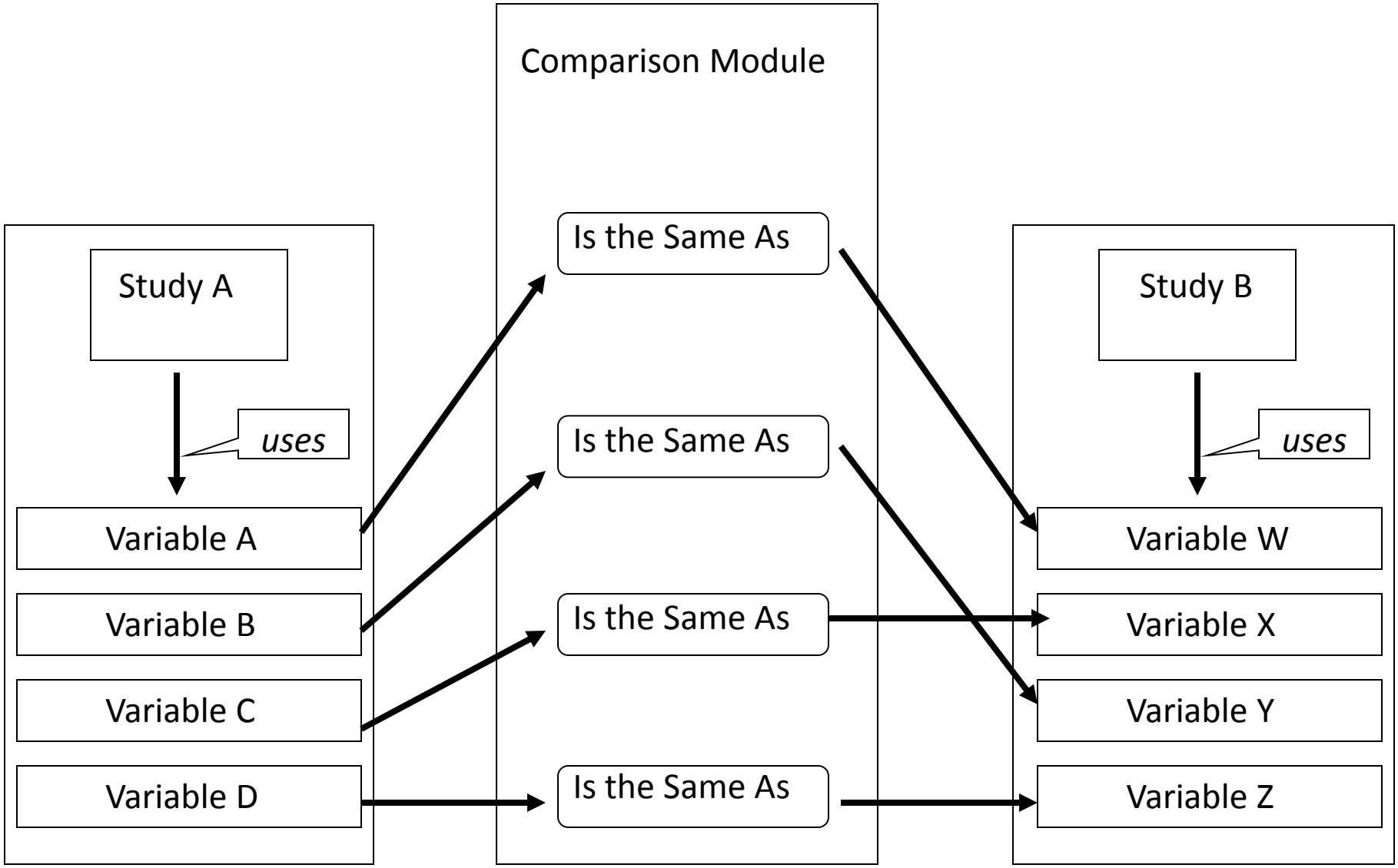
Data Comparison

- To compare data from different studies (or even waves of the same study) we use the *metadata*
  - The metadata explains which things are comparable in data sets
- When we compare two variables, they are comparable if they have the same set of properties
  - They measure the same concept for the same high-level universe, and have the same representation (categories/codes, etc.)
  - For example, two variables measuring “Age” are comparable if they have the same concept (e.g., age at last birthday) for the same top-level universe (i.e., people, as opposed to houses), and express their value using the same representation (i.e., an integer from 0-99)
  - They *may* be comparable if the only difference is their representation (i.e., one uses 5-year age cohorts and the other uses integers) but this requires a *mapping*

DDI Support for Comparison

- For data which is completely the same, DDI provides a way of showing comparability: Grouping
  - These things are comparable “by design”
  - This typically includes longitudinal/repeat cross-sectional studies
- For data which *may* be comparable, DDI allows for a statement of what the comparable metadata items are: the Comparison module
  - The Comparison module provides the mappings between similar items (“ad-hoc” comparison)
  - Mappings are always context-dependent (e.g., they are sufficient for the purposes of particular research, and are only *assertions* about the equivalence of the metadata items)


Study A
uses
Variable A
Variable B
Variable C
Variable D
Study B
uses
Variable A
Variable B
Variable C
Variable X
Group
uses
Variable A
Variable B
Variable C
contains
contains
Study A
uses
Variable D
Study B
uses
Variable X


Comparison Module
Study A
uses
Variable A
Variable B
Variable C
Variable D
Is the Same As
Is the Same As
Is the Same As
Is the Same As
Study B
uses
Variable W
Variable X
Variable Y
Variable Z

DDI 3.0 Modules

- Conceptual Components (concepts, universes)
- Data Collection (survey instruments and collection processing)
- Logical Products (variables, categories, code lists)
- Physical data product (descriptions of file structures)
- Physical Instance (instances of data files)
- Archiving (information about holding, storage, and organizations)
- Comparative (mapping schemes)
- Grouping (for comparison, and longitudinal studies, panels, and series)
- Instance (the wrapper)
- DDI Profile (describes which DDI 3 elements are used)
- Study Unit (describes a single study)

Realizations

- Many different organizations and individuals are involved throughout this process
  - This places an emphasis on versioning and exchange between different systems
- There is potentially a huge amount of metadata reuse throughout an iterative cycle
  - We needed to make the metadata as reusable as possible
- Every organization acts as an “archive” (that is, a maintainer and disseminator) at some point in the lifecycle
  - When we say “archive” in DDI 3.0, it refers to this function

Technical Specifications - Maintainable Schemes

(that's with an 'e' not an 'a')

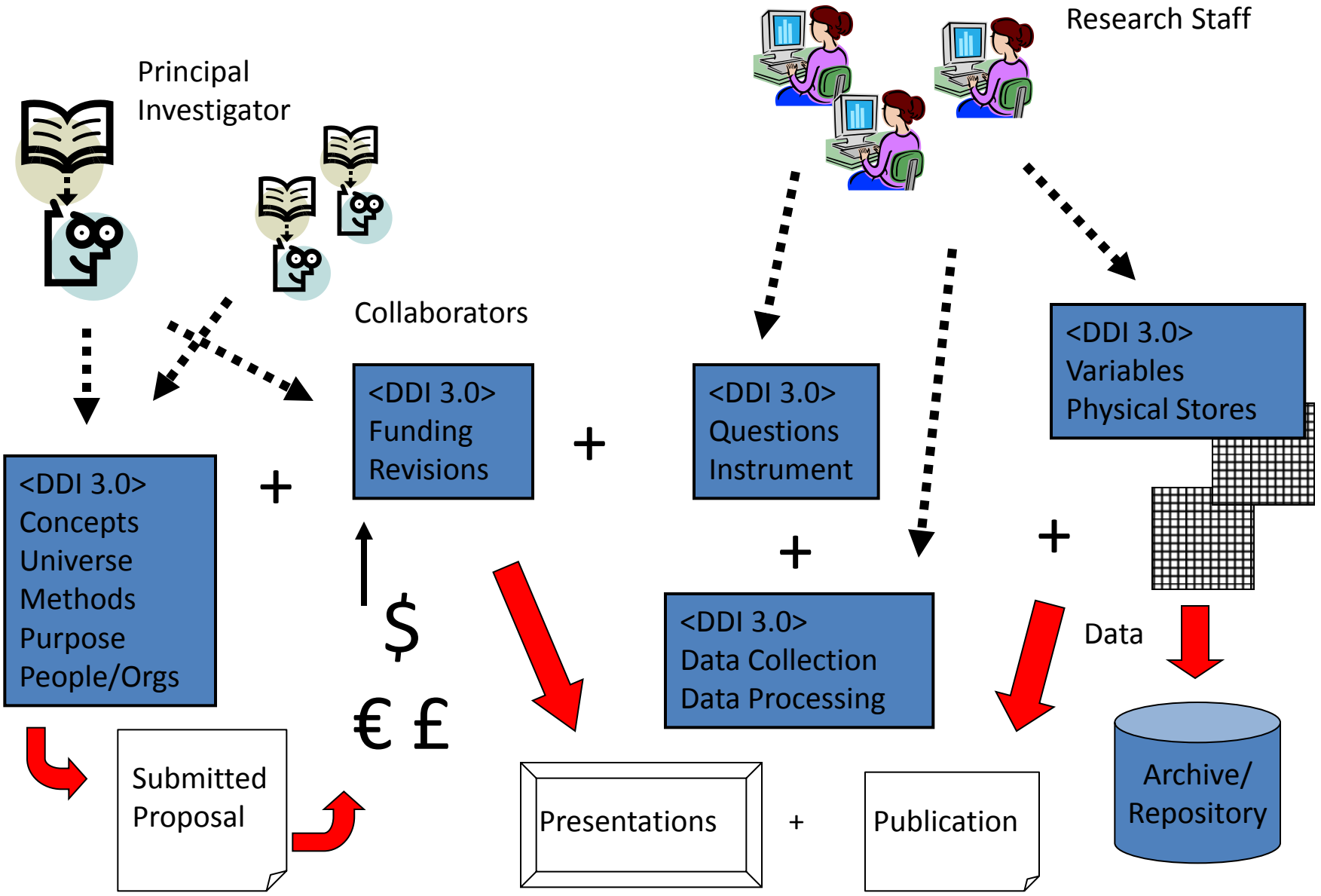
- Category Scheme
- Code Scheme
- Concept Scheme
- Control Construct Scheme
- GeographicStructureScheme
- GeographicLocationScheme
- InterviewerInstructionScheme
- Question Scheme
- NCubeScheme
- Organization Scheme
- Physical Structure Scheme
- Record Layout Scheme
- Universe Scheme
- Variable Scheme

Packages of reusable metadata maintained by a single agency

Technical Specifications – XML Schemas

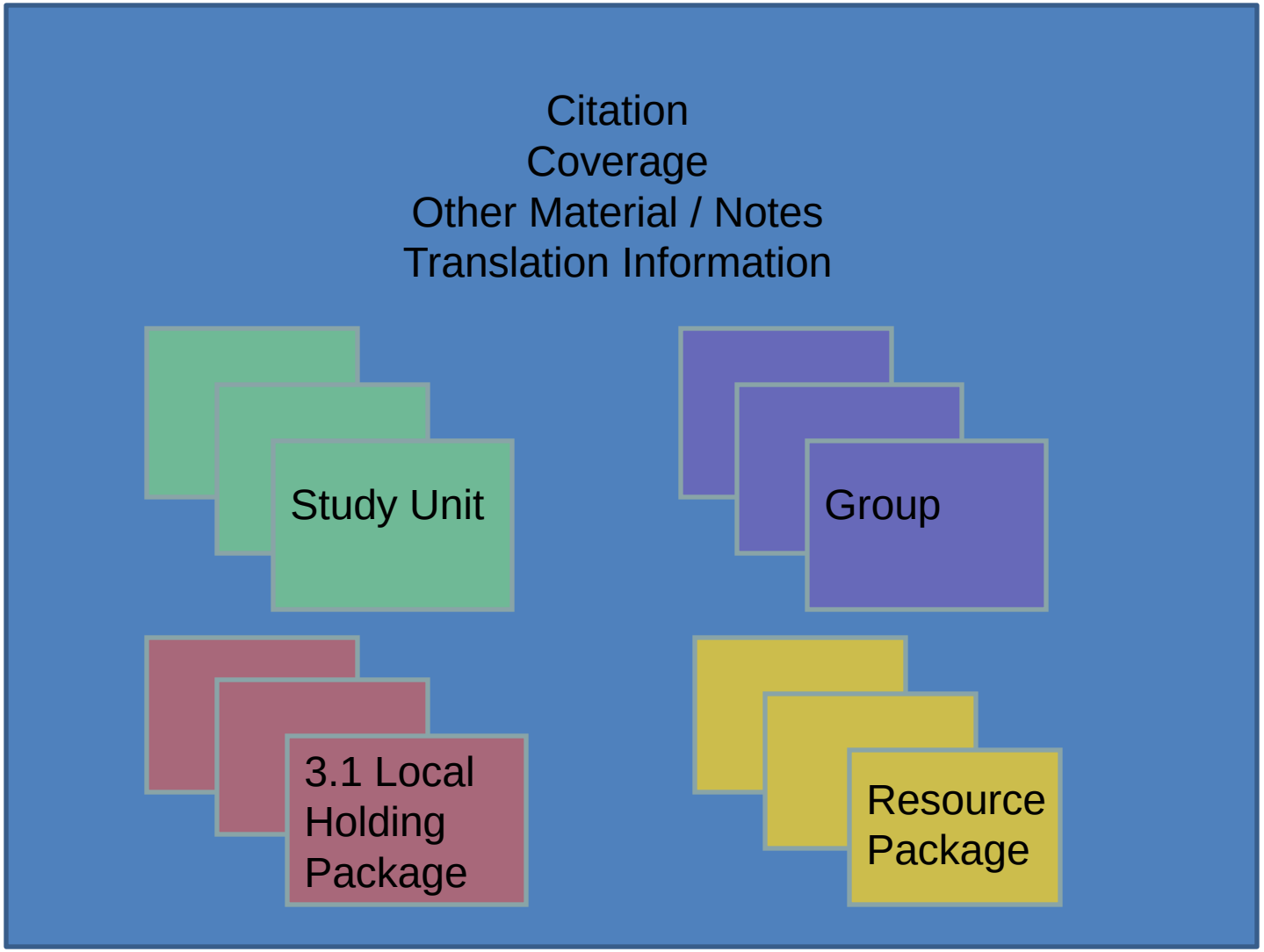
- archive
- comparative
- conceptualcomponent
- datacollection
- dataset
- dcelements
- DDIprofile
- ddi-xhtml11
- ddi-xhtml11-model-1
- ddi-xhtml11-modules-1
- group
- inline_ncube_recordlayout
- instance
- logicalproduct
- ncube_recordlayout
- physicaldataproduct
- physicalinstance
- proprietary_record_layout (beta)
- reusable
- simpledc20021212
- studyunit
- tabular_ncube_recordlayout
- xml
- set of xml schemas to support xhtml

Use within data collection



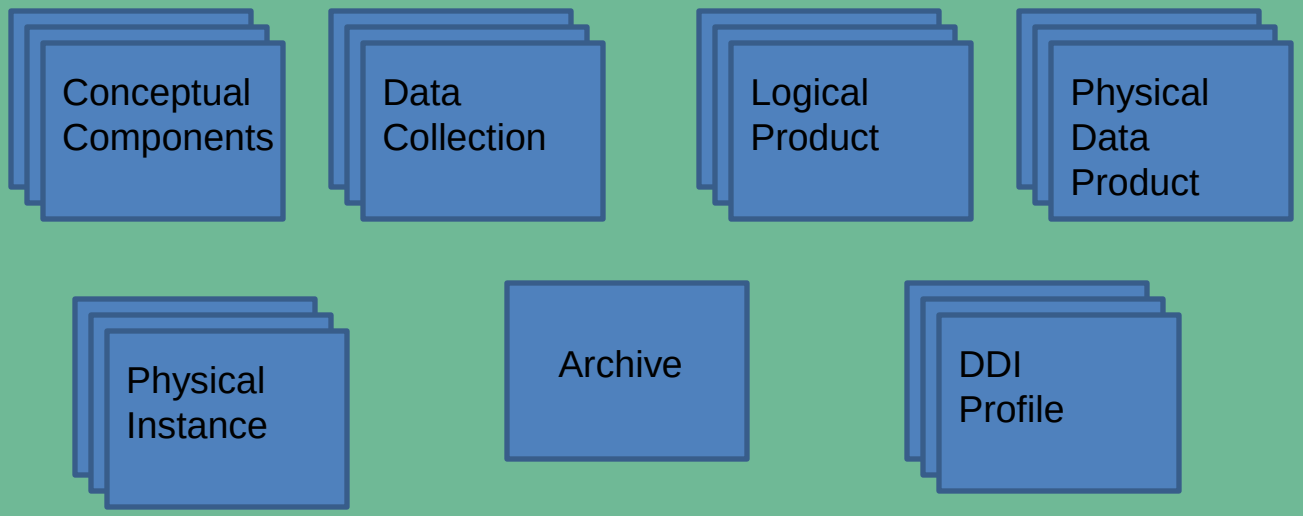
DDI Overall Structure and Component Parts

DDI Instance

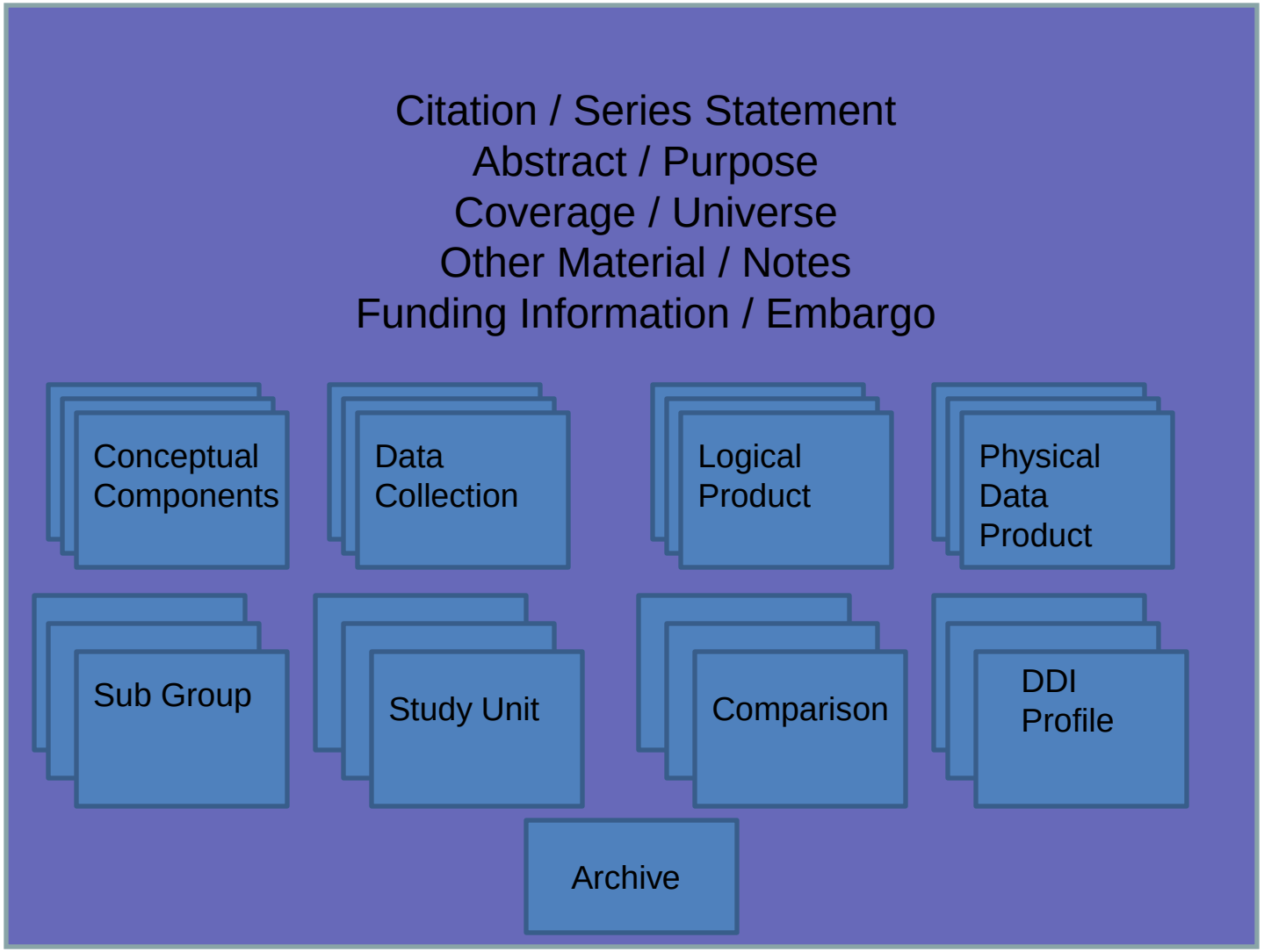


Study Unit

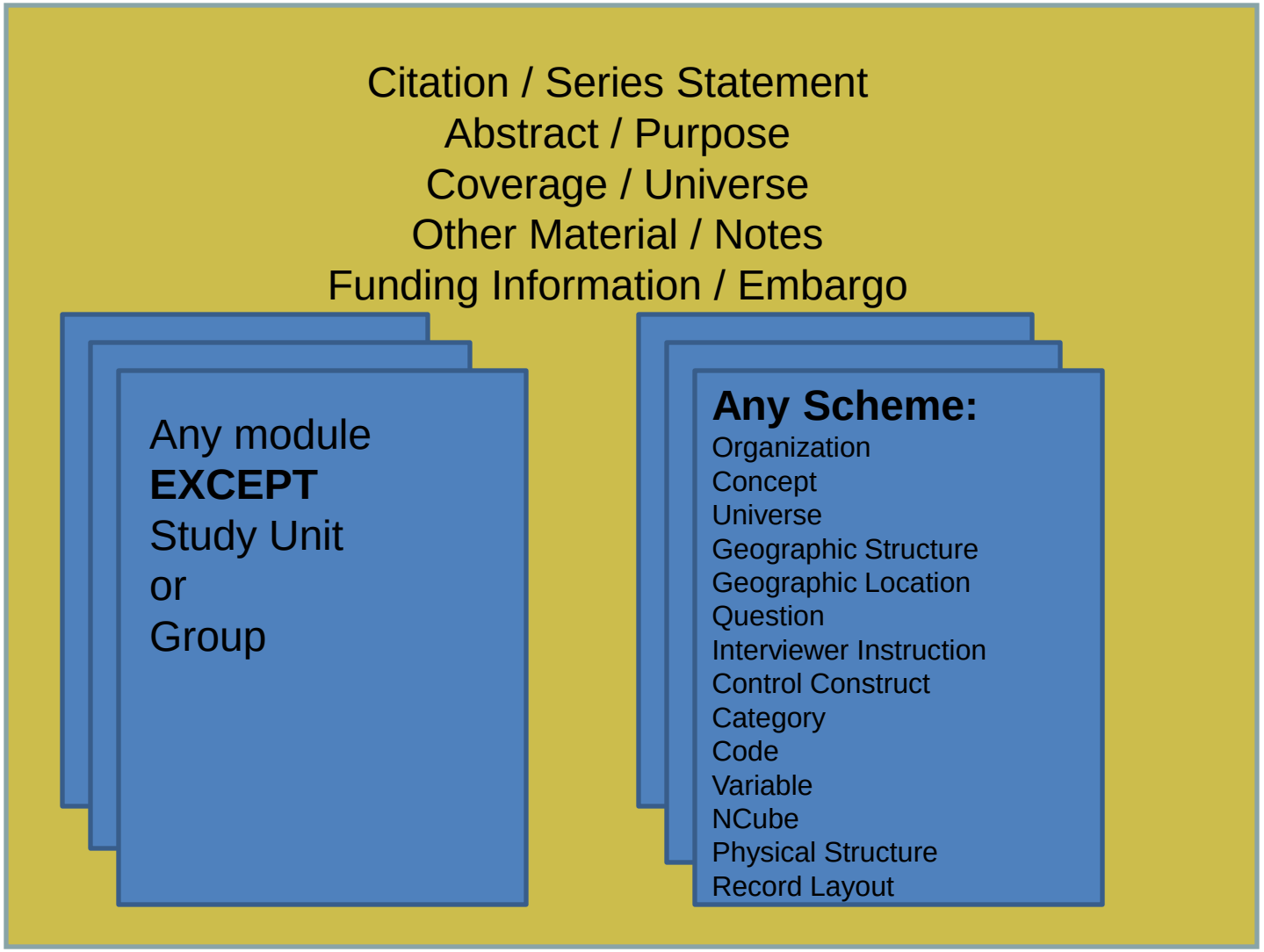
Citation / Series Statement
Abstract / Purpose
Coverage / Universe / Analysis Unit / Kind of Data
Other Material / Notes
Funding Information / Embargo



Group


Citation / Series Statement
Abstract / Purpose
Coverage / Universe
Other Material / Notes
Funding Information / Embargo
Conceptual Components
Data Collection
Logical Product
Physical Data Product
Sub Group
Study Unit
Comparison
DDI Profile
Archive

Resource Package



3.1 Local Holding Package

Citation / Series Statement
Abstract / Purpose
Coverage / Universe
Other Material / Notes
Funding Information / Embargo

Depository Study Unit OR Group Reference:
[A reference to the stored version of the deposited study unit.]

Local Added Content:
[This contains all content available in a Study Unit whose source is the local archive.]

DDI 3 Lifecycle Model and Related Modules

Groups and Resource Packages are a means of publishing any portion or combination of sections of the life cycle

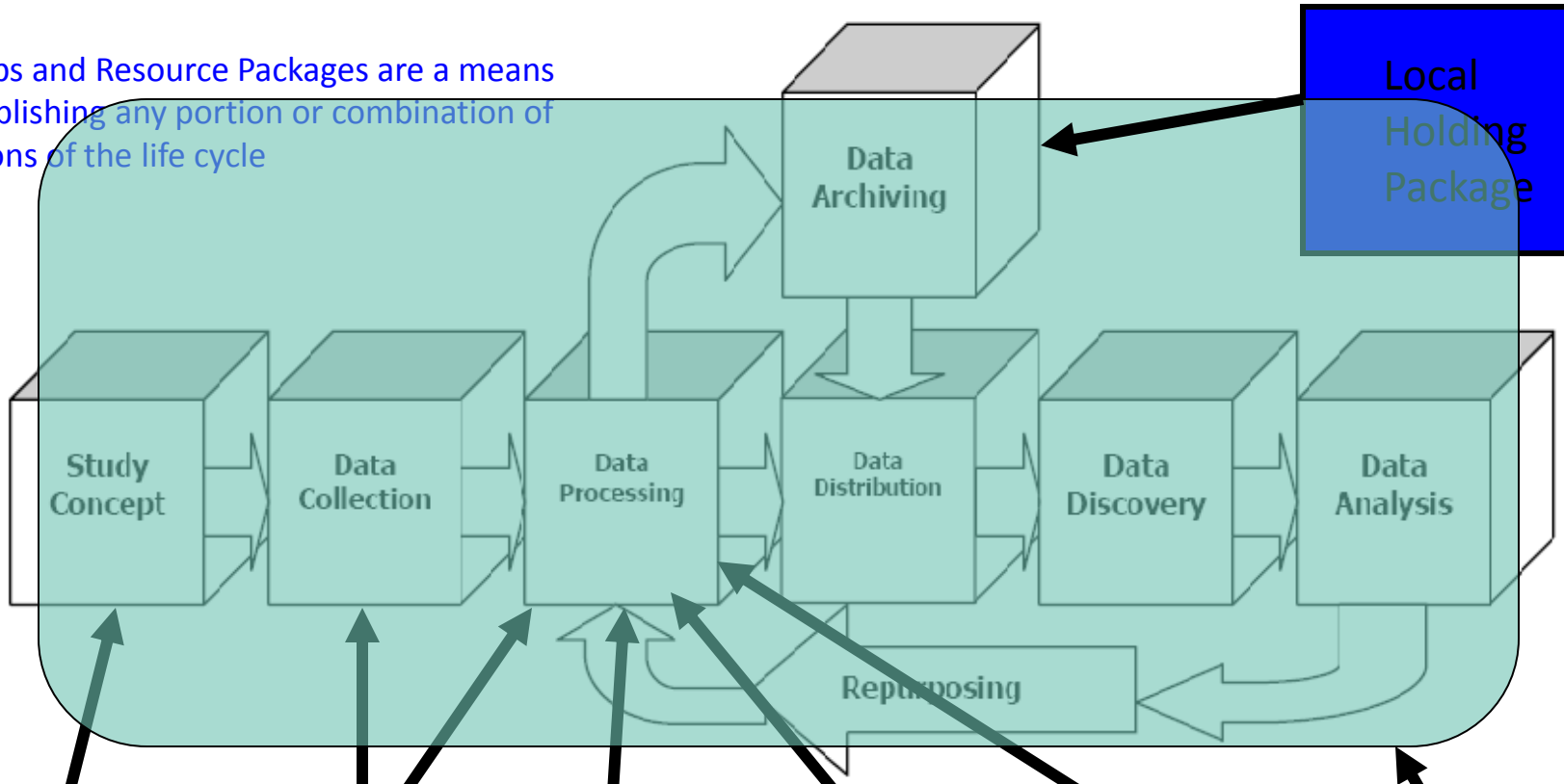


Figure: Combined Life Cycle Model

Study Unit

Data Collection

Logical Product

Physical Data Product

Physical Instance

Archive

Study Unit

- Study Unit
  - Identification
  - Coverage
    - Topical
    - Temporal
    - Spatial
  - Conceptual Components
    - Universe
    - Concept
    - Representation (optional replication)
  - Purpose, Abstract, Proposal, Funding

- Identification is mapped to Dublin Core and basic Dublin Core is included as an option
- Geographic coverage mapped to FGDC / ISO 19115
  - bounding box
  - spatial object
  - polygon description of levels and identifiers
- Universe Scheme, Concept Scheme
  - link of concept, universe, representation through Variable
  - also allows storage as a ISO/IEC 11179 compliant registry

Data Collection

- Methodology
- Question Scheme
  - Question
  - Response domain
- Instrument
  - using Control Construct Scheme
- Coding Instructions
  - question to raw data
  - raw data to public file
- Interviewer Instructions

- Question and Response Domain designed to support question banks
  - Question Scheme is a maintainable object
- Organization and flow of questions into Instrument
  - Used to drive systems like CASES and Blaise
- Coding Instructions
  - Reuse by Questions, Variables, and comparison

Logical Product

- Category Schemes
- Coding Schemes
- Variables
- NCubes
- Variable and NCube Groups
- Data Relationships

- Categories are used as both question response domains and variable representations
- Codes are used as both question response domains and variable representations
- Link representations to concepts and universes through references
- Built from variables (dimensions and attributes)
  - Map directly to SDMX structures
  - More generalized to accommodate legacy data

Physical storage

- Physical Data Structure
  - Links to Data Relationships
  - Links to Variable or NCube Coordinate
  - Description of physical storage structure
    - in-line, fixed, delimited or proprietary
- Physical Instance
  - One-to-one relationship with a data file
  - Coverage constraints
  - Variable and category statistics

Archive

- An archive is whatever organization or individual has current control over the metadata
- Contains persistent lifecycle events
- Contains archive specific information
  - local identification
  - local access constraints

Group

- Resource Package
  - Allows packaging of any maintainable item as a resource item
- Group
  - Up-front design of groups – allows inheritance
  - Ad hoc (“after-the-fact”) groups – explicit comparison using comparison maps for Universe, Concept, Question, Variable, Category, and Code
- Local Holding Package
  - Allows attachment of local information to a deposited study without changing the version of the study unit itself

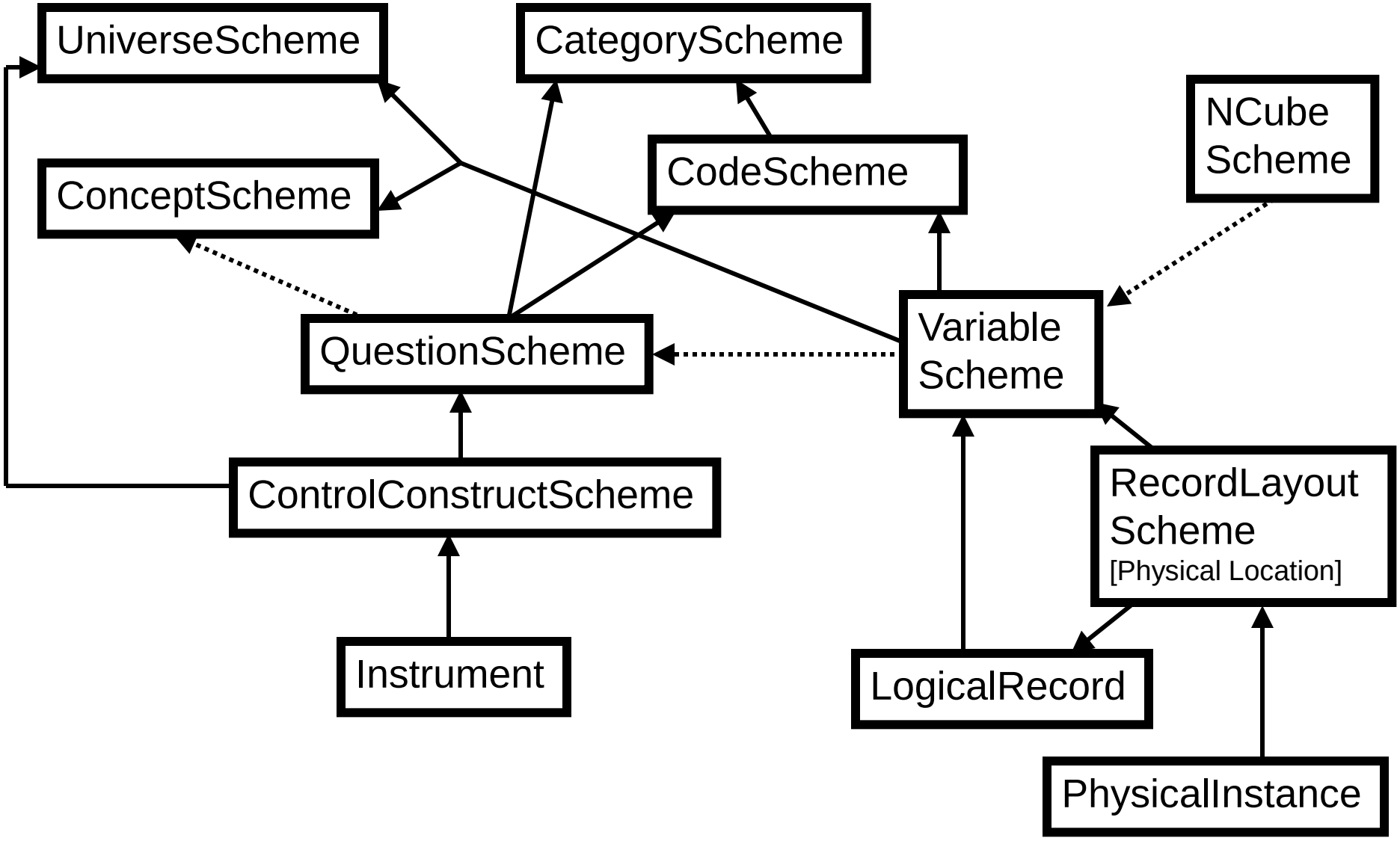
DDI Schemes

- Brief overview of what DDI schemes are and what they are designed to do including:
  - Purpose of DDI Schemes
  - How a DDI Study is built using information held in schemes

DDI Schemes: Purpose

- A maintainable structure that contains a list of versionable things
- Supports registries of information such as concept, question and variable banks that are reused by multiple studies or are used by search systems to location information across a collection of studies
- Supports a structured means of versioning the list
- May be published within Resource Packages or within DDI modules
- Serve as component parts in capturing reusable metadata within the life-cycle of the data

Building from Component Parts



Questionnaires

- Questions
  - Question Text
  - Response Domains
- Statements
  - Pre- Post-question text
  - Routing information
  - Explanatory materials
- Question Flow

Simple Questionnaire

Simple Questionnaire:

1. Sex
   (1) Male
   (2) Female
2. Are you 18 years or older?
   (0) Yes
   (1) No (Go to Question 4)
3. How old are you? _____
4. Who do you live with?

   _____
5. What type of school do you attend?
   (1) Public school
   (2) Private school
   (3) Do not attend school

Simple Questionnaire

Simple Questionnaire:

1. Sex
   (1) Male
   (2) Female
2. Are you 18 years or older?
   (0) Yes
   (1) No (Go to Question 4)
3. How old are you? _____
4. Who do you live with?

   _____
5. What type of school do you attend?
   (1) Public school
   (2) Private school
   (3) Do not attend school

- Questions

Simple Questionnaire

Simple Questionnaire:

1. Sex
   (1) Male
   (2) Female
2. Are you 18 years or older?
   (0) Yes
   (1) No (Go to Question 4)
3. How old are you? _____
4. Who do you live with?

   _____

5. What type of school do you attend?
   (1) Public school
   (2) Private school
   (3) Do not attend school

- Questions
- Response Domains
  - Code
  - Numeric
  - Text

Category and Code Domains

- Use CategoryDomain when NO codes are provided for the category response

  [ ] Yes

  [ ] No

- Use CodeDomain when codes are provided on the questionnaire itself

  1. Yes
  2. No

Category Schemes and Code Schemes

- Use the same structure as variables
- Create the category scheme or schemes first (do not duplicate categories)
- Create the code schemes using the categories
  - A category can be in more than one code scheme
  - A category can have different codes in each code scheme

Numeric and Text Domains

- Numeric Domain provides information on the range of acceptable numbers that can be entered as a response
- Text domains generally indicate the maximum length of the response
- Additional specialized domains such as DateTime are also available

Simple Questionnaire

Simple Questionnaire:

1. Sex
   (1) Male
   (2) Female
2. Are you 18 years or older?
   (0) Yes
   (1) No (Go to Question 4)
3. How old are you? _____
4. Who do you live with?

   _____

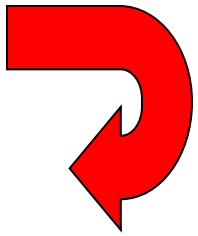
5. What type of school do you attend?
   (1) Public school
   (2) Private school
   (3) Do not attend school

- Questions
- Response Domains
  - Code
  - Numeric
  - Text
- Statements

Simple Questionnaire

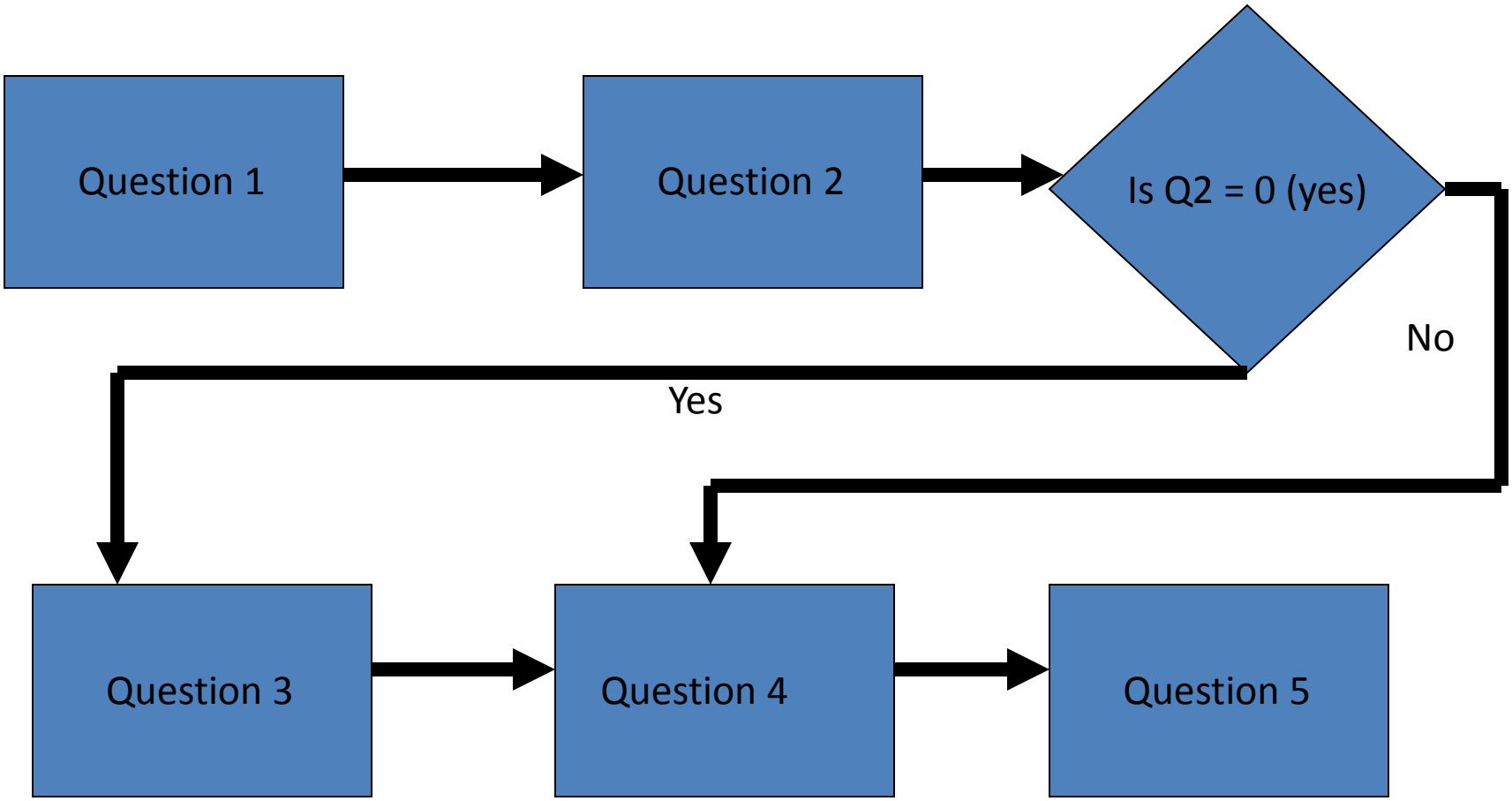
Simple Questionnaire:

1. Sex
   (1) Male
   (2) Female
2. Are you 18 years or older?
   (0) Yes
   (1) No (Go to Question 4)
3. How old are you? _____
4. Who do you live with?
   _____
5. What type of school do you attend?
   (1) Public school
   (2) Private school
   (3) Do not attend school

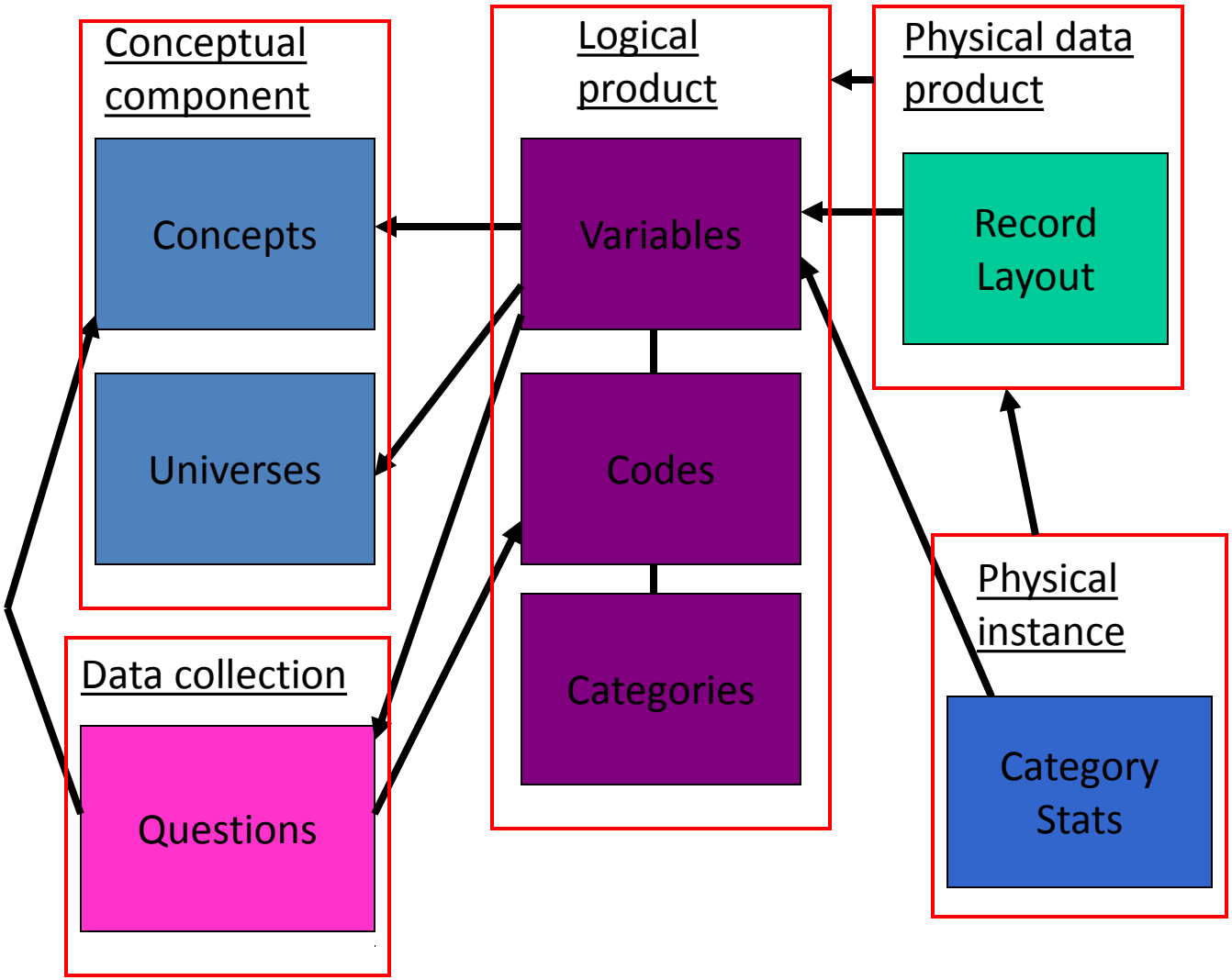


Skip Q3

- Questions
- Response Domains
  - Code
  - Numeric
  - Text
- Statements
- Flow


Question 1
Question 2
Is Q2 = 0 (yes)
No
Yes
Question 3
Question 4
Question 5

DDI 3.0 Modules: Schematic

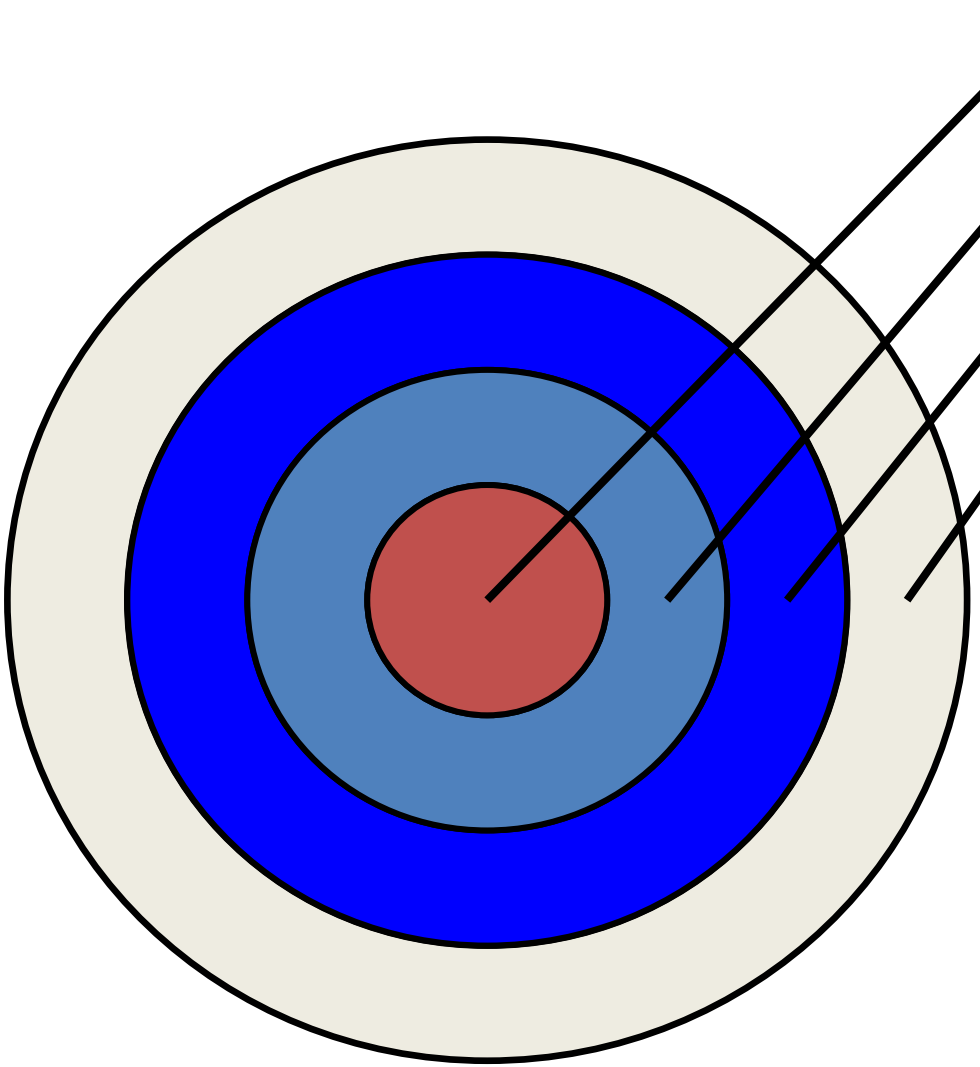


Additional Technical Topics

Maintainable, Versionable, and Identifiable

- DDI 3.0 places and emphasis on re-use
  - This creates *lots* of inclusion by reference!
  - This raises the issue of managing change over time
- The Maintainable, Versionable, and Identifiable scheme in DDI was created to help deal with these issues
- An *identifiable object* is something which can be referenced, because it has an ID
- A *versionable object* is something which can be referenced, and which can change over time – it is assigned a version number
- A *maintainable object* is something which is maintained by a specified agency, and which is versionable and can be referenced – it is given a maintenance agency

Basic Element Types



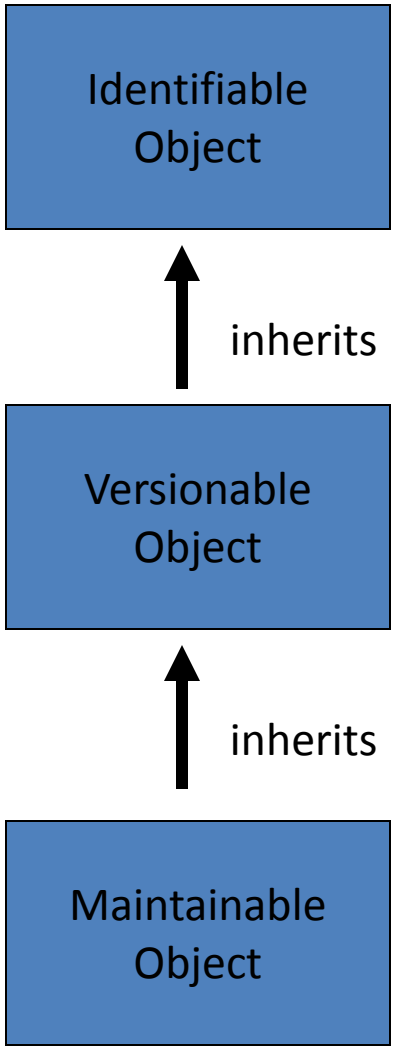
Differences from 2.1

--Every element is NOT identifiable

--Many individual elements or complex elements may be versioned

--A number of complex elements can be separately maintained

In the Model…



Has ID

Eg, Variable, PhysicalRecordSegment

Has ID
Has Version

Eg, Individual, GrossFileStructure, QuestionItem

Has ID
Has Version
Has Agency

Eg, VariableScheme, QuestionScheme, PhysicalDataProduct

What Does This Mean?

- As different pieces of metadata move through the lifecycle, they will change.
  - At a high level, “maintainable” objects represent packages of re-usable metadata passing from one organization to another
  - Versionable objects represent things which change as they are reviewed within an organization or along the lifecycle
  - Identifiable things represent metadata which is reused at a granular level, typically within maintainable packages
- The high-level documentation lists out all maintainables, versionables, and identifiables in a table

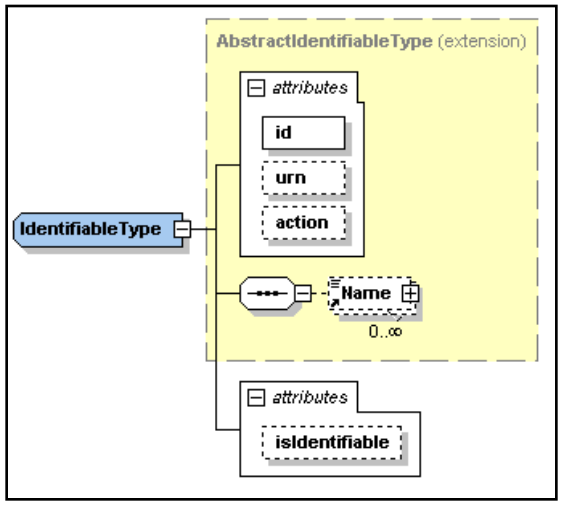
Inheritance of Agency and Version

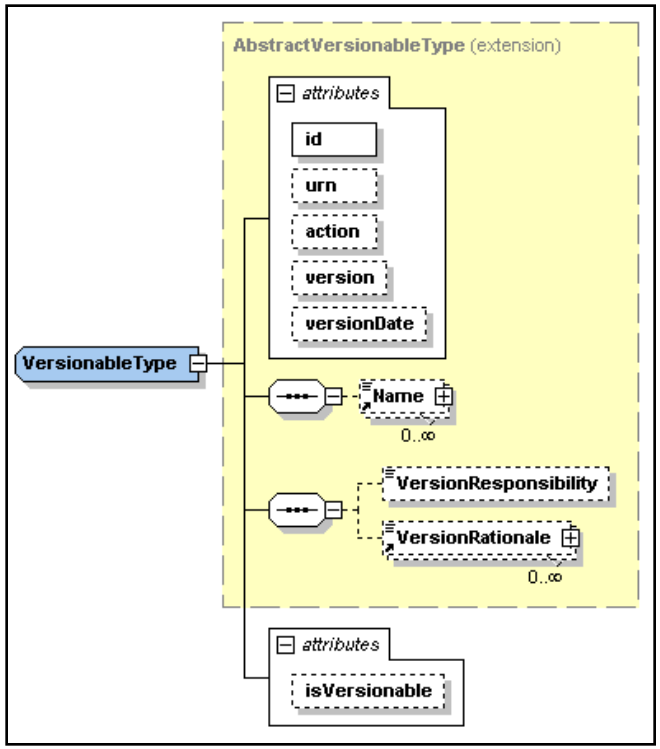
- In DDI 3.0 XML instances, identifiables and versionables live in maintainable schemes or modules
  - All of the children of the scheme inherit that scheme's agency
  - If identifiables live inside of a versionable, the identifiables inherit the version number of the versionable
- All of these objects always *implicitly* have an agency, a version, and an ID
- This becomes clear in the way DDI 3.0 identifiers are structured

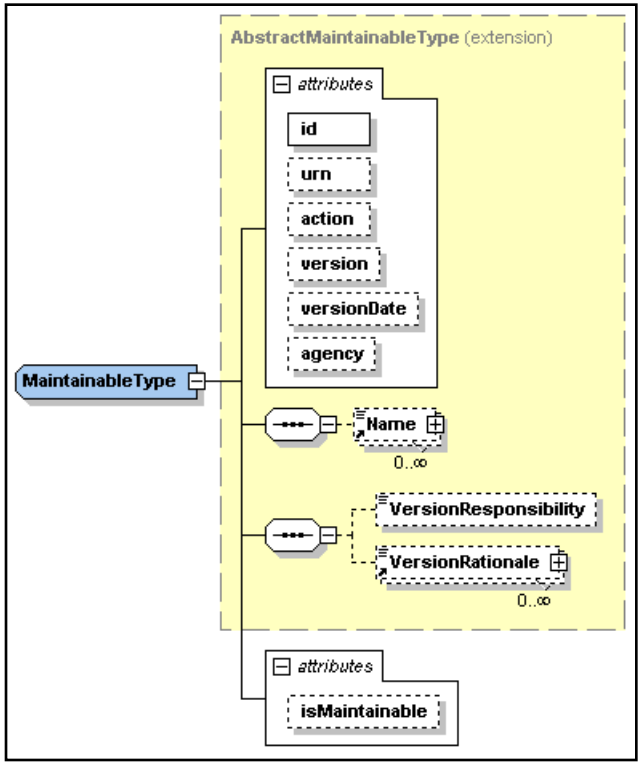
DDI 3.0 Identifiers

- There are two ways to provide identification for a DDI 3.0 object:
  - Using a set of XML fields
  - Using a specially-structured URN
- The structured URN approach is preferred
  - URNs are a very common way of assigning a universal, public identifier to information on the Internet
  - However, they require explicit statement of agency, version, and ID information in DDI 3.0
- Providing element fields in DDI 3.0 allows for much information to be defaulted
  - Agency can be inherited from parent element
  - Version can be inherited or defaulted to “1.0”

Identification Types







Parts of the Identification Series

- Identifiable Element
  - Identifier:
    - ID
    - Identifying Agency
    - Version
    - Version Date
    - Version Responsibility
    - Version Rationale
- Variable
  - Identifier:
    - V1
    - pop.umn.edu
    - 1.1 [default is 1.0]
    - 2007-02-10
    - Wendy Thomas
    - Spelling correction

DDI Identifiers: Elements

- Typical appearance (identifiable):

```
<pdp:DataItem @id="AB347" isIdentifiable="true">
    …
</pdp:DataItem>
```

- Typical appearance (versionable):

```
<lp:Variable id="V101" version="1.1" versionDate="2007-02-12"
    isVersionable="true">
    <r:VersionResponsibility>Wendy Thomas</r:VersionResponsibility>
    <r:VersionRationale>Spelling Correction</r:VersionRationale>
    …
</lp:Variable >
```

- Typical appearance (maintainable):

```
<lp:VariableScheme id="STUDY012345_VarSch01" agency ="pop.umd.edu"
    version="1.0" isMaintainable="true">
    …
</dc:Identifier>
```

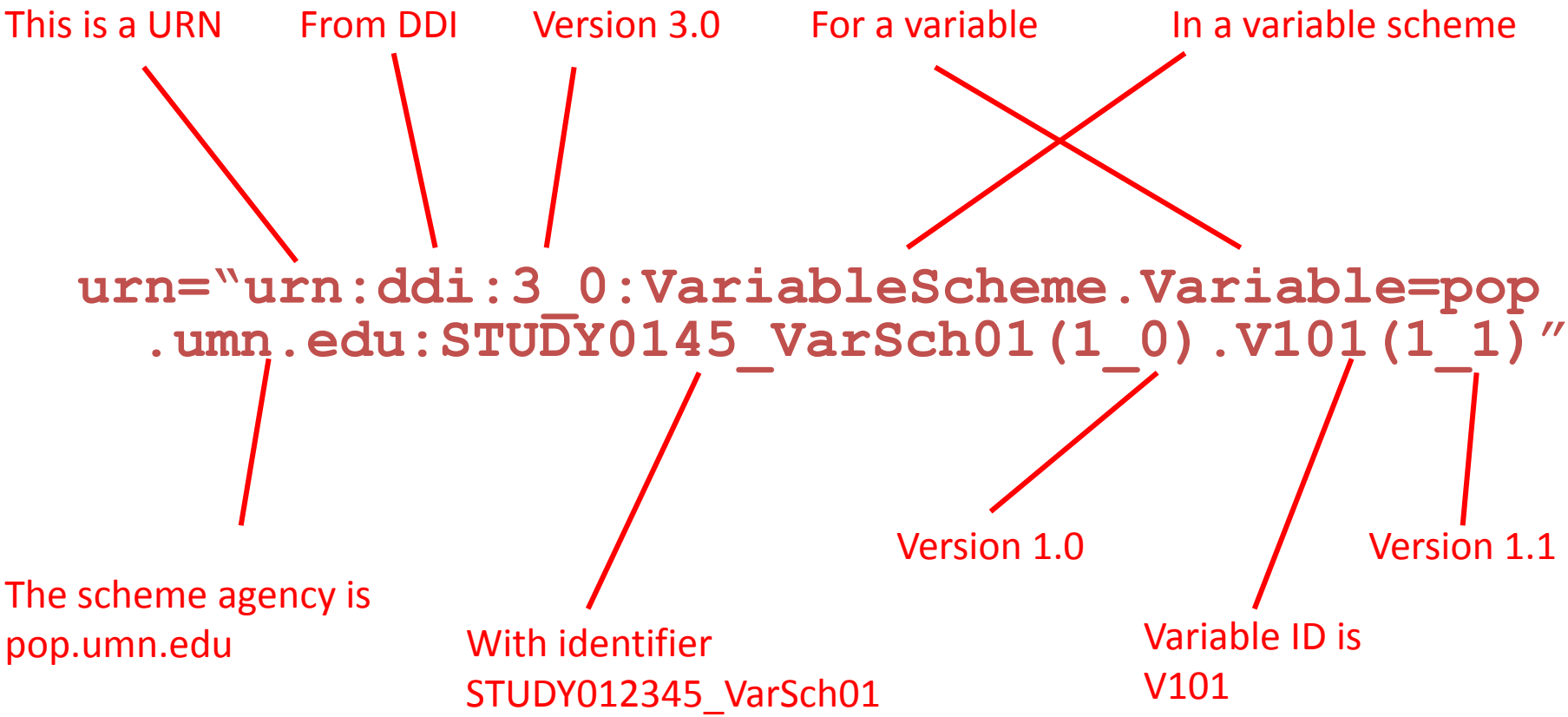
- Note that version and agency may be defaulted/inherited, which means they do not need to be supplied in the local element
  - In a simple example, they are given once for the whole study
  - The object type is determined by the containing element

The URN

```
urn="urn:ddi:3_0:VariableScheme.Variable=pop.
  umn.edu:STUDY0145_VarSch01(1_0).V101(1_1)"
```

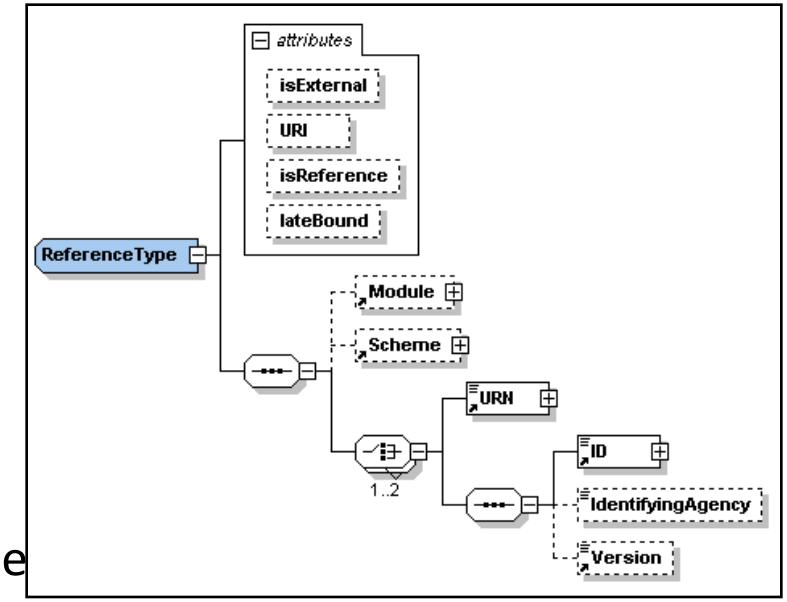
- Declares that its a ddi version 3.0 element
- Tells the type of element it is
- Gives the identifying agency
- Provides its unique ID
  - Note that this includes both a maintainable ID and element ID as uniqueness must be maintained within a maintainable object rather than within the agency
- There are generic tools for resolving URNs
  - They are mapped to local URLs

URN Detailed Example



DDI Internal References

- References in DDI may be within a single instance or across instances
  - Metadata can be re-packaged into many different groups and instances
- Identifiers must provide:
  - The containing module (optional)
    - Agency, ID, and Version
  - The containing maintainable (a scheme
    - Agency, ID, and Version
  - The identifiable/versionable object within the scheme
    - ID (and version if versionable)
- Like identifiers, DDI references may be using URNs or using element fields



Extensions to DDI

- DDI allows for extensions to the standard, using a mechanism based on XML Schema
  - Extending party declares their own namespace
  - Imports the DDI schemas to be extended
  - Declares new elements and types
  - Attaches the new elements to existing imported elements using xs:extends
  - Attaches the extended elements (in the extension namespace) using substitution groups (with extended DDI element as the head of the group)

Effect of Extensions

- Native DDI elements are still in the DDI namespace
- New elements are in the declared extending namespace
  - These can be detected as added extensions
- Extended elements are in the extending namespace
  - But they are true extensions of native types in the related DDI namespace

DDI 3 Resources, Events, and Tools

DDI Resources

- DDI Alliance Site
  - http://www.ddialliance.org
  - General link to all resources/news
  - Link to Sourceforge for standards distributions
  - Link to prototype page – good for examples
- Tools/Resources Page
  - http://tools.ddialliance.org
  - Best place for tools, slides, and resources
- Mailing Lists
  - www.icpsr.umich.edu/mailman/admin/
  - All of the lists starting with "DDI" are related to DDI topics
    - DDI Users (best place to link into the group)
    - List for each sub-committee (not all groups are active)

DDI Resources (cont.)

- Open Data Foundation Site
  - **www.opendatafoundation.org**
  - White papers, other resources/tools
- DDI Developer's Group
  - **groups.google.com/group/ddi-developers**
  - A DDI developer's group has recently been started on Google Groups.
  - This is an unofficial group, but is a good place to get answers for questions about technical DDI issues

DDI Resources (cont.)

- DDI Agency Registry
  - http://tools.ddialliance.org/?lvl1=community&lvl2=agencyid
  - Sign up for unique global agency identifier – helps provide interoperability between organizations
  - Currently taking pre-registrations – site will be permanent in future
- International Household Survey Network
  - http://surveynetwork.org
  - DDI 2-based toolkit available for developing countries (some free tools)
  - Catalog of surveys, many documented in DDI

Recent DDI Publications

- Best Practices Across the Data Life Cycle
  - www.ddialliance.org/resources/publications/working/bestpractices
  - The work of 25 individuals who came together at Schloss Dagstuhl, in Wadern Germany, in November 2008
- Use Cases
  - www.ddialliance.org/resources/publications/working/usecases
  - These papers on DDI 3 use cases are the outcomes of a workshop held at Schloss Dagstuhl - Leibniz Center for Informatics in Wadern, Germany, November 2-6, 2009.
- IASSIST Quarterly v.33:1-2 spring/summer 2009
  - www.iassistdata.org/iq/
  - A special double feature focusing on various projects related to DDI 3 and it's enhanced features
  - Articles related to DDI can be found in many issues of the IQ

Best Practices

- Implementation and Governance
- Work flows - Data Discovery and Dissemination: User Perspective
- Work flows - Archival Ingest and Metadata Enhancement
- Work flows for Metadata Creation Regarding Recoding, Aggregation and Other Data Processing Activities
- Controlled Vocabularies
- Creating a DDI Profile
- DDI 3.0 Schemes
- Versioning and Publication
- DDI as Content for Registries
- Management of DDI 3.0 Unique Identifiers
- DDI 3.0 URNs and Entity Resolution
- High-Level Architectural Model for DDI Applications

Use Cases

- Questasy: Documenting and Disseminating Longitudinal Data Online Using DDI 3
- Building a Modular DDI 3 Editor
- Using DDI 3 for Comparison
- Extracting Metadata From the Data Analysis Workflow
- Questionnaire Management and DDI: The QDDS Case
- Grouping of Survey Series Using DDI 3
- An Archive's Perspective on DDI 3

DDI Events

- IASSIST
  - www.iassistdata.org
  - Not an official DDI event, but many DDI-related presentations and meetings
  - DDI Alliance Expert Committee meets before or after every year
  - 37th Meeting in Vancouver BC Canada, 30 May – 3 June 2011
  - DDI Workshops given day before the meeting
  - Annual meetings go US-Canada-US-Outside North America-US-Canada-US-Outside North America etc.

DDI Events (cont.)

- European DDI User's Group
  - 2st meeting will be in Utrecht, The Netherlands 8-9 December 2010
  - Preceded by a DDI Implementers 2.5 day workshop, 6-8 December 2010
- GESIS-Sponsored Autumn Events
  - Schloss Dagstuhl workshops
- Open Data Foundation meetings
  - Spring meeting in Europe
  - Winter meeting in the US
  - DDI is a major topic of discussion, especially as it is used with other standards

Tools/Projects

- DDI 3 has only been an official standard since April 2008
  - Despite this, many tools are being developed
  - Some useful tools already exist
- Some tools are available, others are projects which would be willing to share code (or partner) as the basis for further development
  - The list may not be complete
  - IASSIST has a DDI Tools panel every year – see online presentations

Tools/Projects (cont.)

- Nesstar (developed by Norwegian Social Sciences Data Services)
  - Commercial product supporting DDI 1/2
  - Looking now at support for DDI 3
  - Provides an editing interface, visualization/tabulation, and server-to-server data exchange
  - Nesstar editor is used by the IHSN Metadata Toolkit, which adds publishing functionality for HTML, PDF, and CD-ROMs
  - Useful for migration to 3

Tools/Projects (cont.)

- DDI Foundation Tools Program
  - Joint initiative by several organizations to develop open-source tools for DDI 3
  - Includes DeXT (UKDA) and GESIS-developed tools for transformations to and from DDI 1/2 – 3.0 and statistical packages (SAS, SPSS)
  - Provides a utilities package for Java development, including validation, XML beans, URN resolution
  - Now developing a suite of tools for editing DDI 3 instances based on a common application framework (work is lead out of the Danish Data Archive)

Tools/Projects (cont.)

- Canadian RDC Network
  - Producing DDI 3-based tools for many DDI use cases
    - Editing
    - Migration from DDI 1/2
    - Registries
    - Repositories
    - Metadata mining
  - All tools will be open-source when completed (over next 2 years)

Tools/Projects (cont.)

- Colectica (by Algenta)
  - Commercial tool supporting survey instrument creation, and other editing functions of DDI
  - Has a repository component
  - Supports DDI 1/2, 3, Blaise, Cases, and CSPro files
  - DDI 3.1 is the native file format
- CSPro
  - Is currently developing support for DDI 3
  - Already supports DDI 1/2
  - Free product for conducting surveys (similar to CASES and Blaise)

Tools/Projects (Cont.)

- Space-Time Research
  - Adding DDI 1/2 and 3 support to their line of products (SuperCross, SuperWeb, etc.), for loading micro-data into their proprietary databases
  - Commercial tool providing point-and-click functionality for tabulation of microdata
  - Support for SDMX expression of tabulations
    - Uses SDMX RESTful Web services
- Questasy
  - Based on an online documentation tool for the LISS panel study at CentERdata
  - Willing to partner to productize the code base
  - Database-driven application using PHP and other easy Web development technologies

Tools/Projects (Cont.)

- Exanda
  - Online tabulation system based on DDI 3
  - Intended to be released as open source, but no committed delivery date
  - Uses freely available software components (Flex, Apache Cocoon, etc.)
- QDDS
  - Documentation system for questionnaires developed by GESIS, Leibniz Institute for the Social Sciences
  - Uses DDI 1/2, plans for supporting DDI 3 in future
  - Freely available, but not open source

Questions?