DDI and Its Role in Modernizing Official Statistics

Joachim Wackerow, Chair Scientific Board, DDI Alliance

Sixth Session of the Statistical Commission of
Organisation of Islamic Cooperation
Konya, Turkey, November 5-6, 2016

Today's Presentation

- What is DDI?
- DDI Specifications with example use in official statistics
- DDI and reference specifications in official statistics
- DDI Alliance

What is DDI?

- International standard for describing data from the social, behavioral, and economic sciences and other data on human activity
- Main focus of DDI is to capture metadata about microdata
- Metadata for most of the processes of statistical production
- A metadata specification of and for the community
- Freely available

Benefits

- Exchange and preservation in system-independent format with support for any natural language (XML Schema)
- Make documentation a byproduct of the process
- Decrease redundant work
- Reusability of key survey components
- Increased data harmonization potential
- Rich content (over 1000 items in DDI-Lifecycle)
- Metadata re-use across the life cycle (DDI-Lifecycle)
- Data management and curation
- Support for longitudinal data and comparison (DDI-Lifecycle)
- Support for a global network (unique identifier in DDI-Lifecycle)

Development

- Established in 1995 by an international group
- Creation of a specification for the content, structure, and exchange of machine-readable data dictionaries
- Expanded to document and manage data resources across the data life cycle
- Aligned with standards such as Dublin Core, ISO/IEC 11179 (metadata registry), and GSIM

Overview of Specifications

- Two major development lines:
  - DDI-Codebook
  - DDI-Lifecycle
- Additional specifications:
  - Controlled Vocabularies, recommended sets of terms and definitions for selected DDI elements
  - RDF Vocabularies/Ontologies
    - DDI-RDF Discovery
    - XKOS
- In development:
  - DDI 4 - next generation, model-driven approach

DDI-Codebook

- For data dissemination, data dictionary for a single study
- After-the-fact description
- Current version 2.5.1
- Example users
  - CESSDA - Consortium of European Social Science Data Archives, soon in the form of ERIC (European Research Infrastructure Consortium)
  - CRDCN - Canadian Research Data Centre Network
  - IHSN - International Household Survey Network
  - TurkStat - Turkish Statistical Institute, Harzemli
- Major software: Nesstar (metadata/data publishing)

IHSN - International Household Survey Network

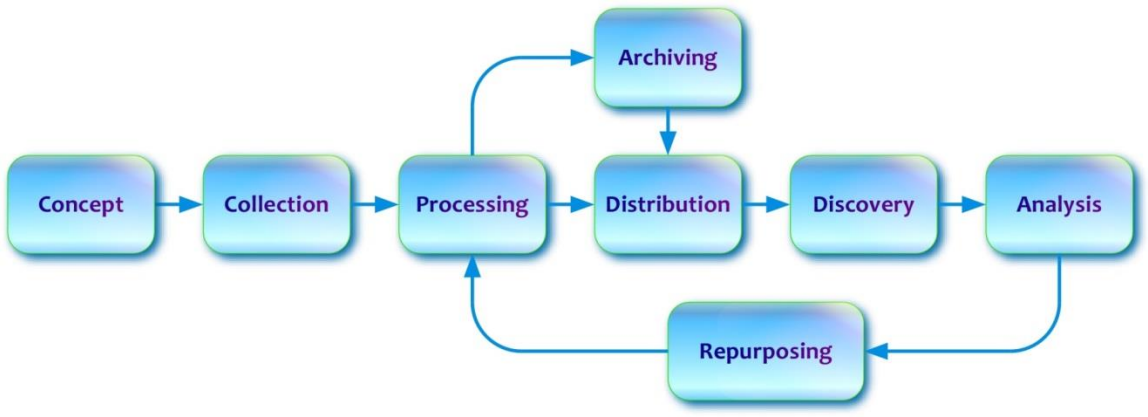
- Mission
  - improve the availability, accessibility, and quality of survey data within developing countries
  - encourage the analysis and use of this data by national and international development decision makers, the research community, and other stakeholders
- IHSN Metadata Editor (Nesstar Publisher), compliant with the DDI 2.* and the Dublin Core metadata standards
- NADA - Microdata Cataloging Tool

International Household Survey Catalog

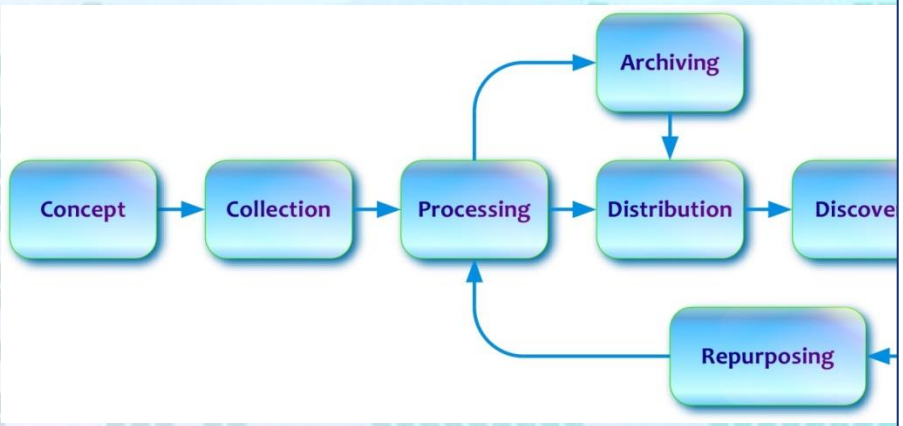
DDI-Lifecycle

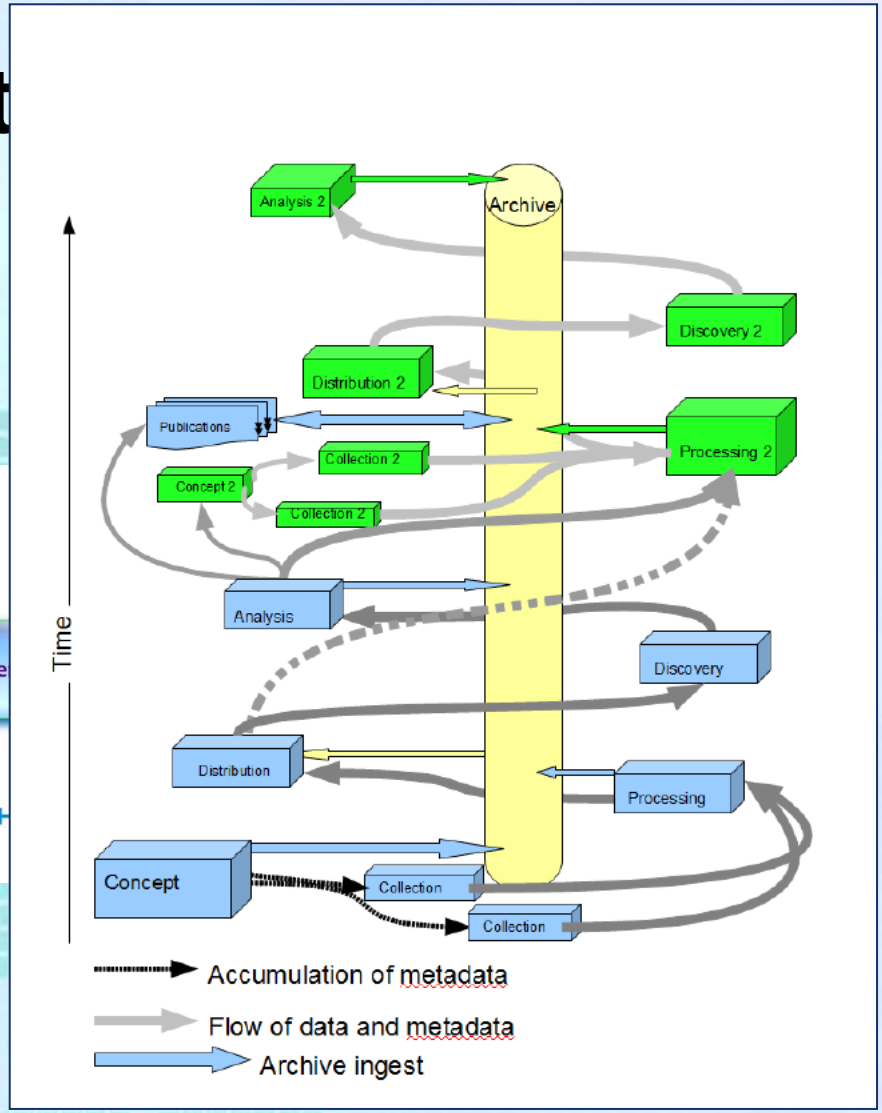
- Focus on metadata re-use
  - Coverage of a single study along the data lifecycle
  - Several waves of data collection
  - Several versions of data and metadata as they change across the data lifecycle
  - Support for a global network. Unique identifier for major DDI items.
- Ad-hoc collection of datasets for purposes of comparison
- Metadata in DDI-Codebook can als be described using DDI-Lifecycle
- Current version 3.2
- Major software: Colectica - design, document, and publish statistical data and survey research using open data standards

Data Lifecycle Orientation I

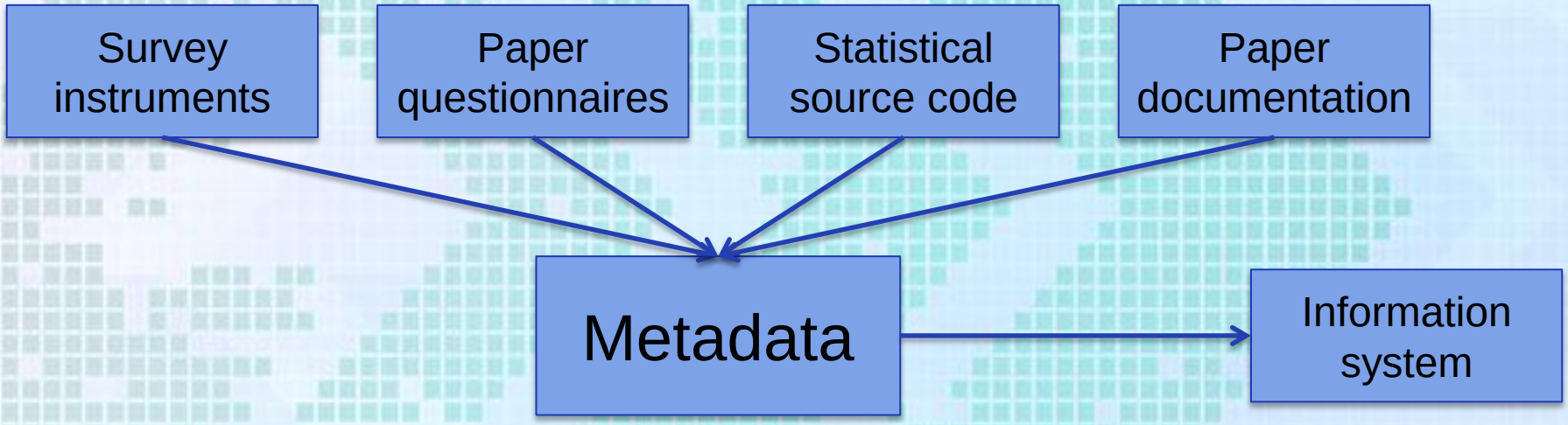


Data Lifecycle Orient

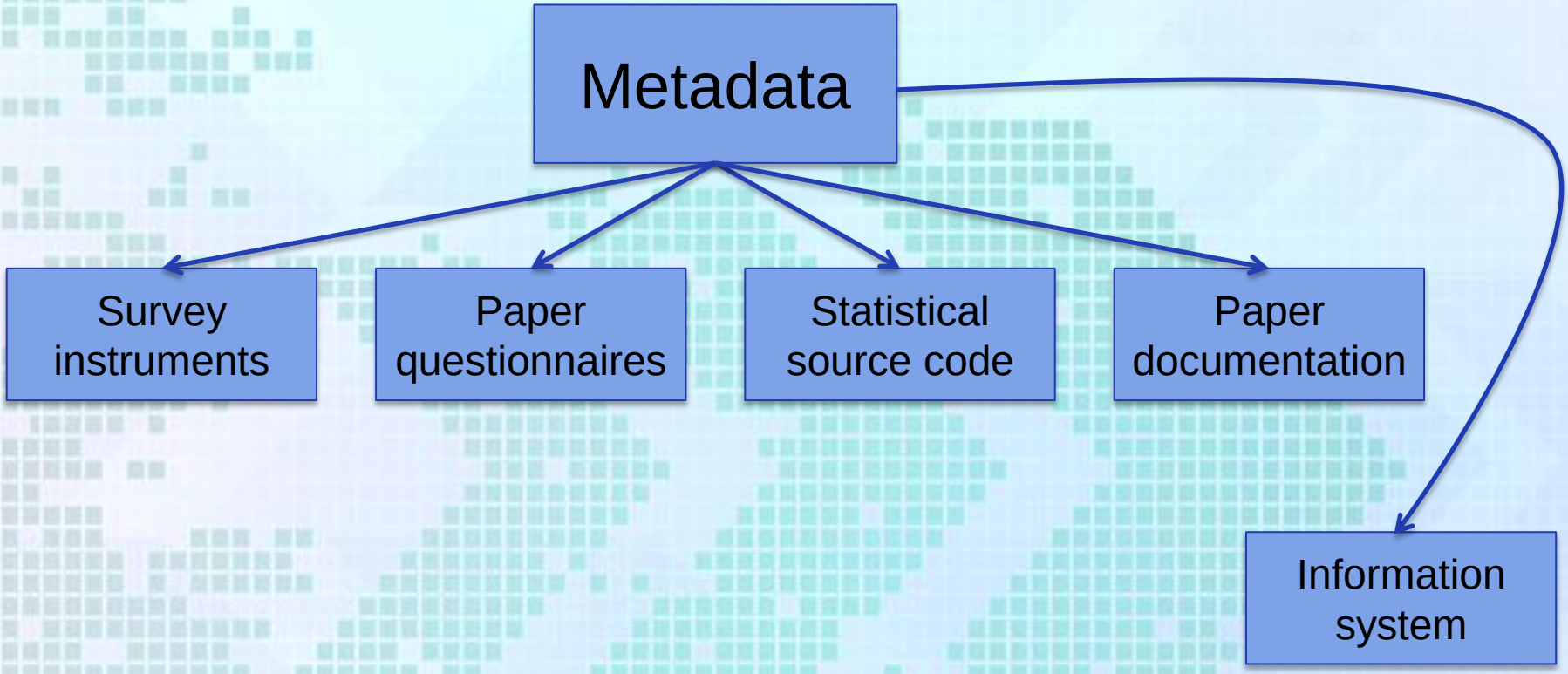




Convential Creation of Metadata



DDI-Lifecycle Supports Metadata Driven Approach



Example Use of DDI-Lifecycle in NSIs

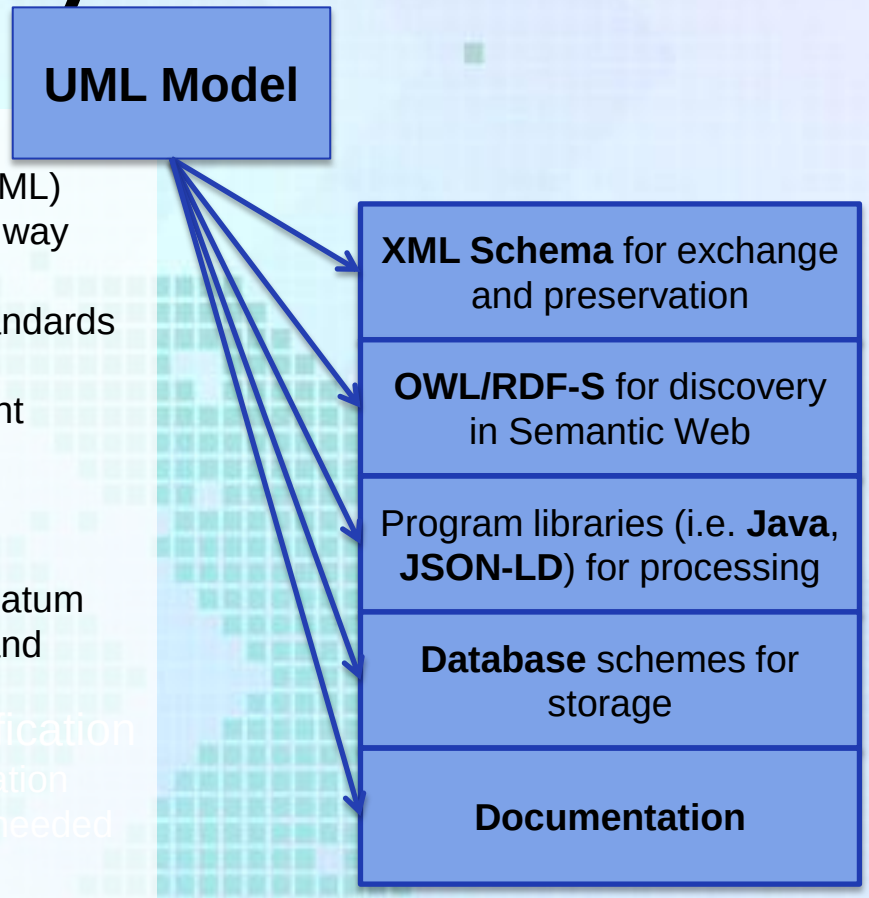
- INSEE - French National Institute of Statistics and Economic Studies
  - Questionnaire generation
- Statistics Netherlands
  - Blaise - computer-assisted interviewing (CAI) system and survey processing tool, export to DDI
- Statistics Denmark
  - DDI-based system for handling concepts and quality information
- Statistics New Zealand
  - Metadata repository with Colectica

RDF Vocabularies

- DDI-RDF Discovery
  - The vocabulary leverages the DDI specification to create a simplified version for the discovery of microdata sets and related metadata in the Semantic Web
  - It is based on a subset of the DDI XML formats of DDI-Codebook and DDI-Lifecycle
  - **Existing DDI XML instances** can be transformed into this RDF format and therefore **exposed in the Web of Linked Data**
- XKOS - Extended Knowledge Organization System
  - XKOS leverages the Simple Knowledge Organization System (SKOS) for managing statistical classifications and concept management systems, since SKOS is widely used
  - XKOS extends SKOS for the needs of statistical classifications
  - Example application: INSEE uses XKOS

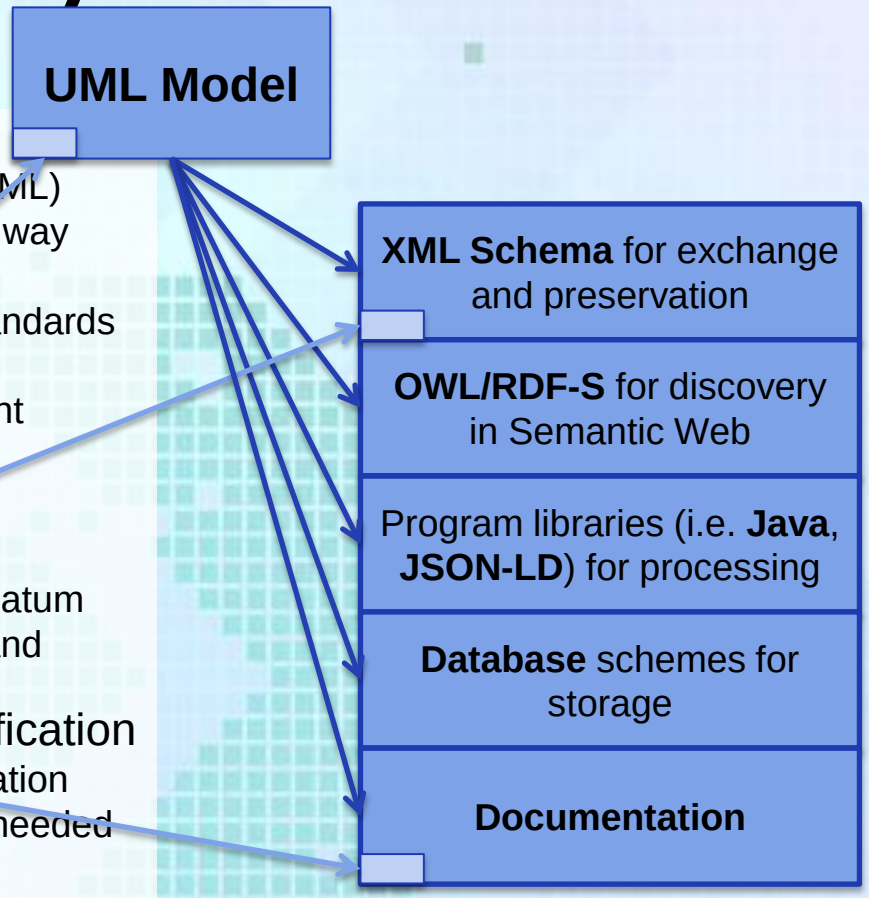
DDI 4 (in development)

- Model-based and model-driven approach
  - Specification in Unified Modeling Language (UML)
  - Easier to develop and maintain in a consistent way
  - Easier to understand
  - Better interaction with other disciplines and standards
  - Provides multiple representations/bindings
  - Will enable more efficient software development
- Additional content, i.e.:
  - Process model
  - Abstraction of data capture/collection/source
  - Universal data description based on „atomic“ datum
  - Sampling, survey implementation, weighting, and paradata
- Functional Views provide subsets of specification
  - Support specific user perspectives on specification
  - Provide easier understanding and limit use to needed parts



DDI 4 (in development)

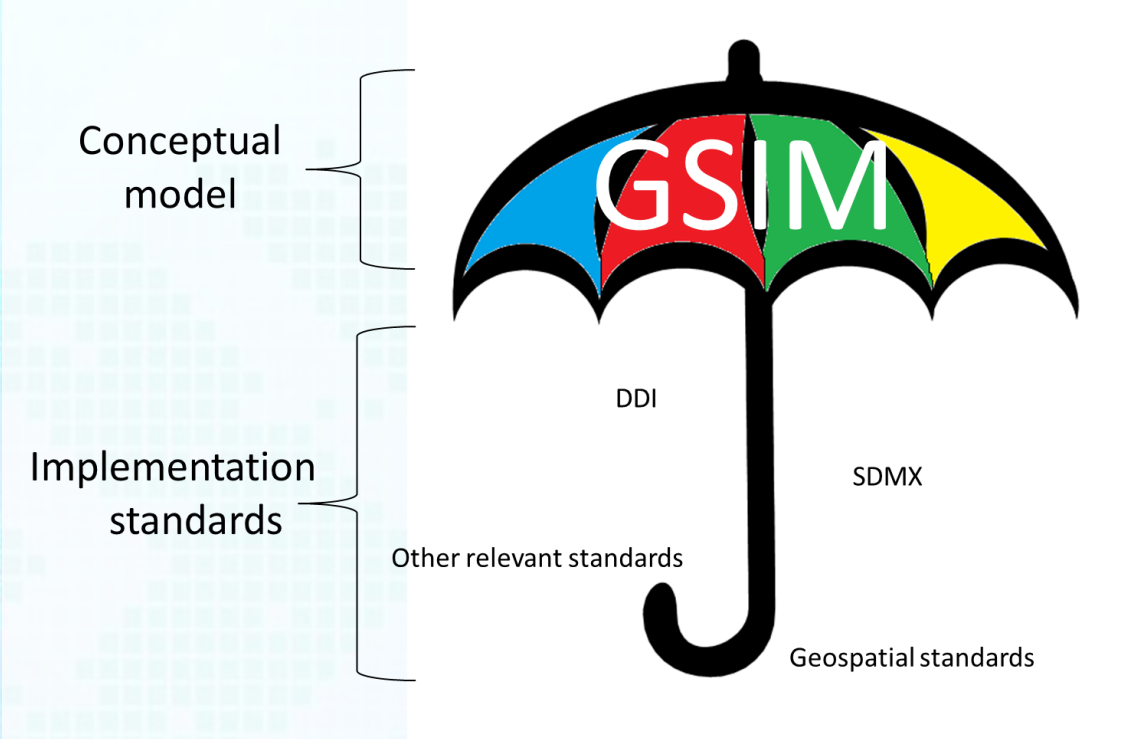
- Model-based and model-driven approach
  - Specification in Unified Modeling Language (UML)
  - Easier to develop and maintain in a consistent way
  - Easier to understand
  - Better interaction with other disciplines and standards
  - Provides multiple representations/bindings
  - Will enable more efficient software development
- Additional content, i.e.:
  - Process model
  - Abstraction of data capture/collection/source
  - Universal data description based on „atomic“ datum
  - Sampling, survey implementation, weighting, and paradata
- Functional Views provide subsets of specification
  - Support specific user perspectives on specification
  - Provide easier understanding and limit use to needed parts



Interoperability within and across Organizations – GSIM and DDI

Interoperability increases if following is used:

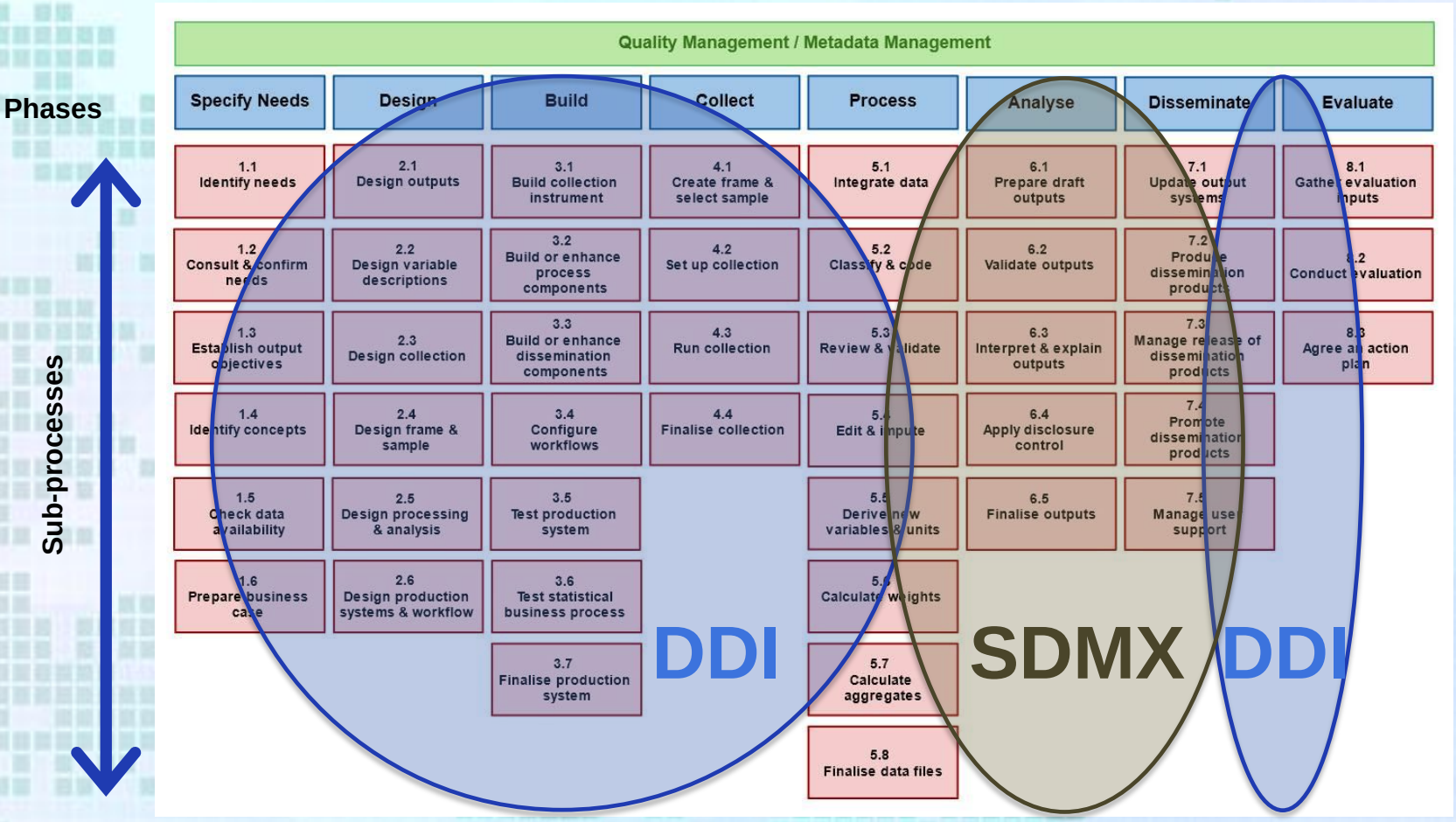
- GSIM conceptual model
- Relevant GSIM implementation standards like DDI
- DDI Profiles (subset of valid DDI objects used by an agency for a specified purpose)



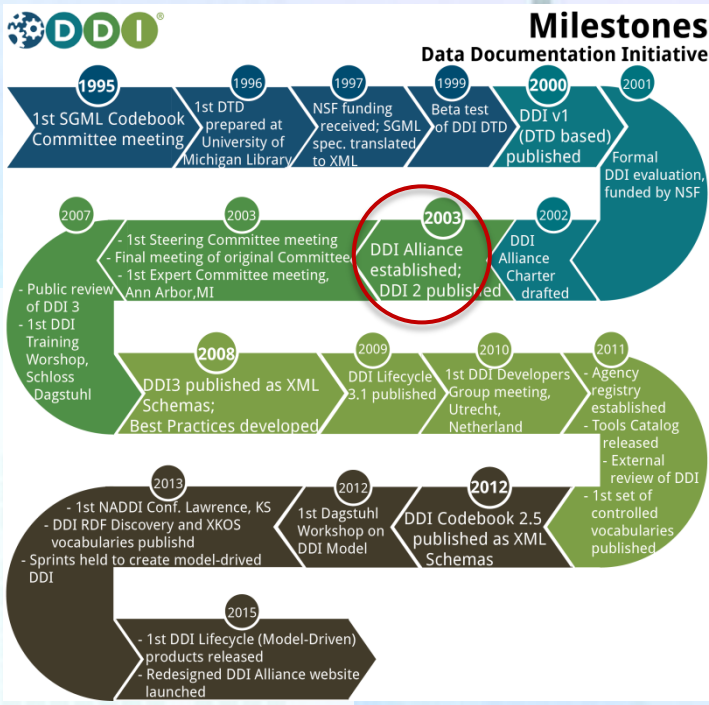
GSIM and DDI - Mutual Influences

- GSIM
  - Conceptual model
  - Reference framework of information objects
  - Common language related to the data and metadata used throughout the statistical production process
  - Complimentary relationship with DDI and SDMX
  - Information objects can be mapped to DDI and SDMX, or own system
- DDI and DDI
  - DDI influenced GSIM, i.e. variable cascade and questionnaire
  - DDI 4 (in development) uses ideas from GSIM and makes improvements, i.e. process model, „atomic“ datum, variable cascade, abstraction of data source

Use of DDI and SDMX for GSBPM



DDI Alliance



- Self-sustaining membership alliance
- 42 institutional members
- Membership includes
  - data archives, data producers, data distributors, research centers, national statistical institutes, and software companies
  - 5 National Statistical Institutes: Australia, Denmark, France, Netherlands, and New Zealand
  - 2 supranational organizations: Eurostat and World Bank
- Members have a voice in shaping and developing the specifications

Thank you for your attention

Questions?

www.ddialliance.org

joachim.wackerow@gesis.org