STANDARDS BASED METADATA USAGE AT STATISTICS DENMARK AND STATISTICS NEW ZEALAND

Colectica FedCASIC 2014

Outline

1. Introduction to DDI Lifecycle standard
2. DDI Lifecycle usage at
    1. Statistics New Zealand
    2. INSEE
    3. Statistics Denmark
3. Next Steps

What is DDI Lifecycle?

- DDI Lifecycle is a metadata **standard**
- Expressed using XML
- DDI provides a **vocabulary** for describing surveys, questions, and data

DDI at a Glance

Study

- Series
- Study Waves
- Quality

Survey

- Data Collection
- Instrument
- Control Flow
- Question

Data

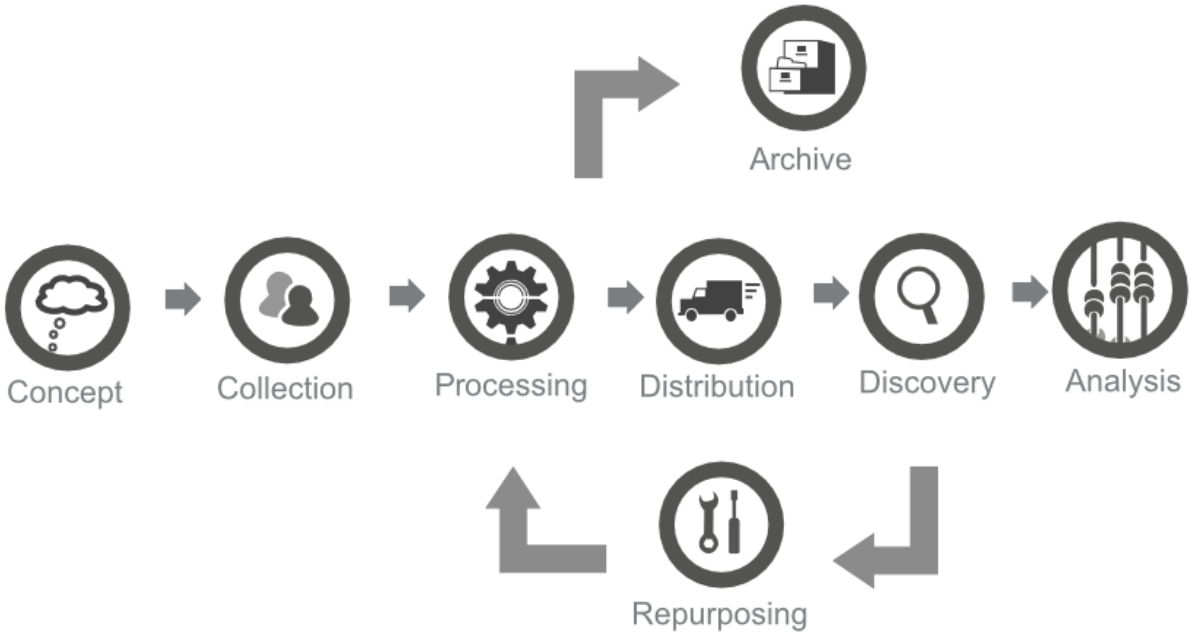
- Datasets
- Data Relations
- Variables

Foundational

- Concept
- Universe
- Organization
- Code Lists
- Category Lists
- Category

Why DDI?

- Document the full data lifecycle in a standard manner



Metadata Banks

- DDI 3 supports the concept of metadata registries
  - Question banks
  - Variable banks
  - Code lists, concept definitions, or anything else

Use Cases:
DDI and Official Statistics

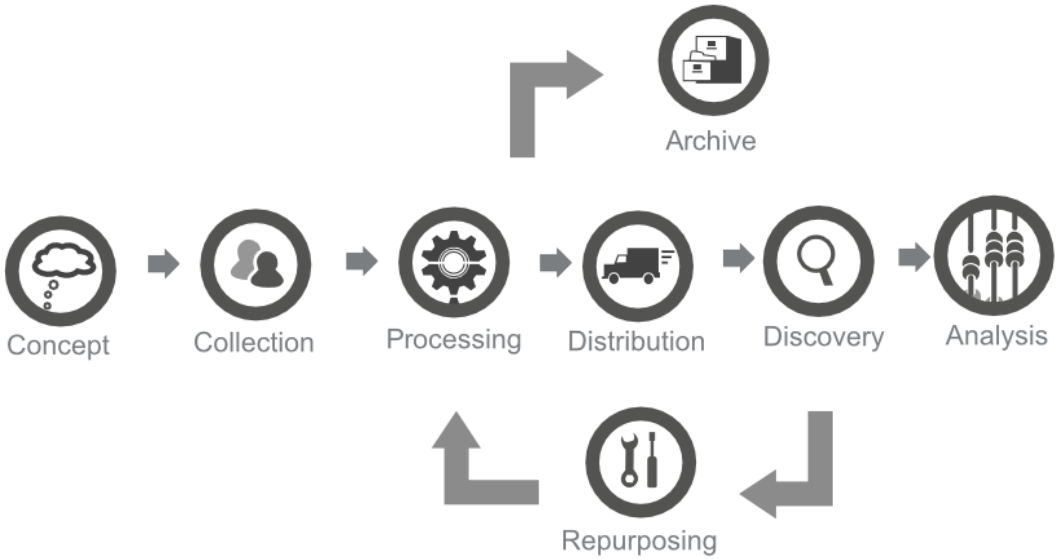
DDI Users

- Official Statistics
  - Statistics New Zealand
  - INSEE
  - Statistics Denmark
- Long-term Longitudinal Studies
  - National Children's Study (NIH and BAH)
  - Midlife in the United States
  - Wisconsin Longitudinal Study
- Archives
  - UCLA Social Science Data Archive
  - Yale Institution for Social and Policy Studies

Adopting DDI in Official Statistics

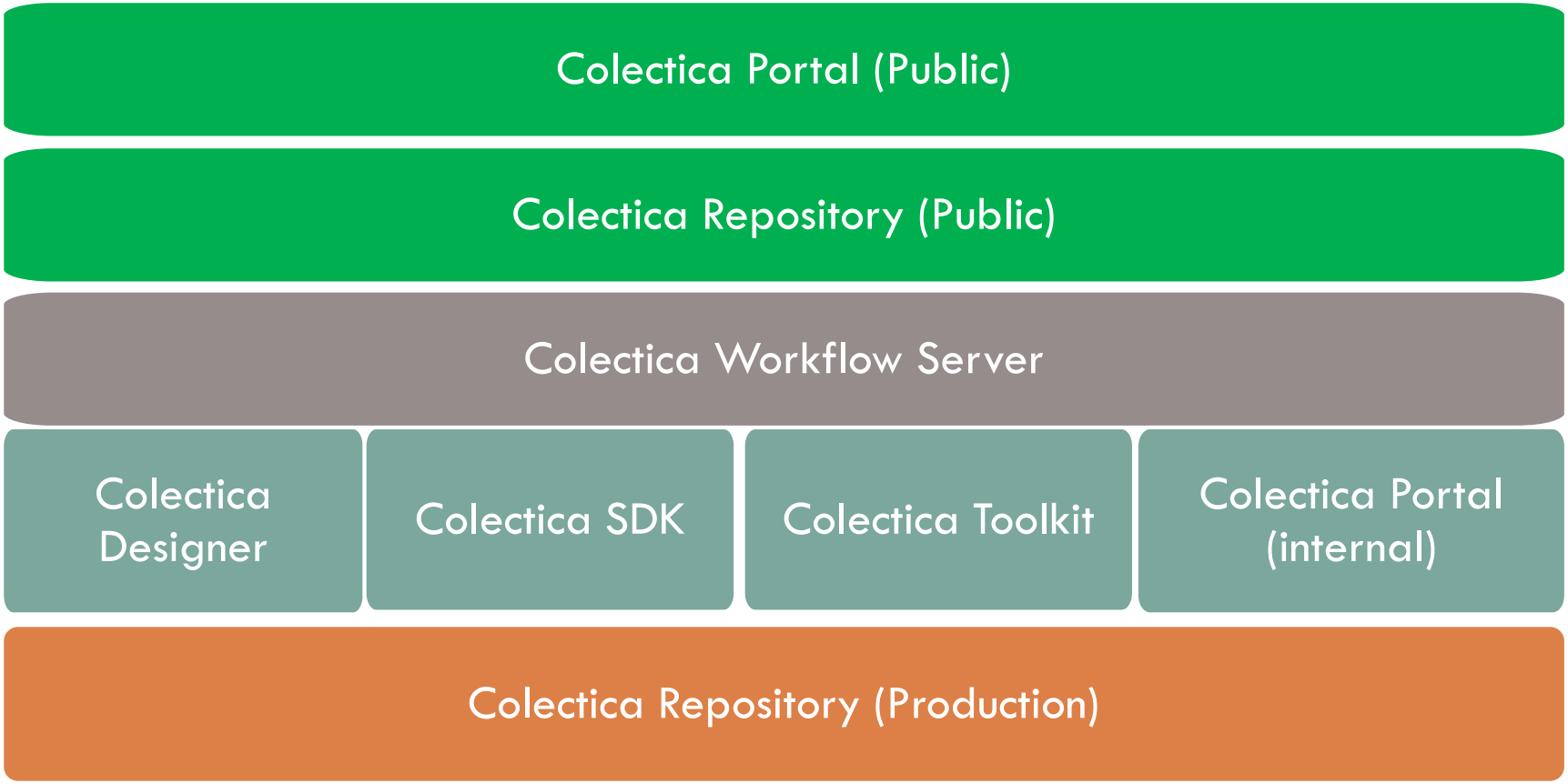
- "Industrialize" the production of statistics
- Doing more with fewer resources
- Improving quality
- Creating a common vocabulary across institutions

Statistics New Zealand
Metadata Infrastructure Project



- Create a single, canonical source for all this information
- Solution: central repository

Architecture: Repository



Key Result 1 – Metadata Capture

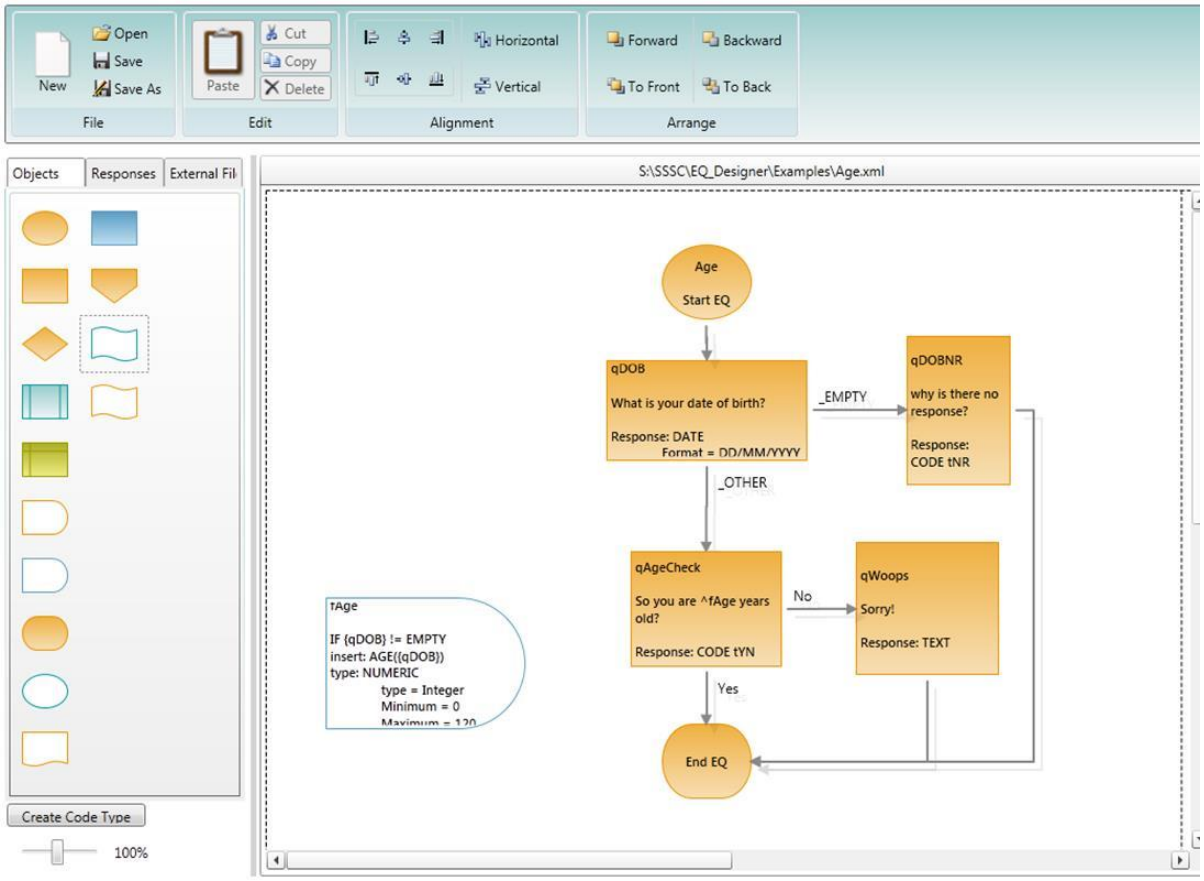
"We used to record all metadata at the end of the lifecycle."

"Now, curators capture the information when they first think of it."

Key Result 2 - Archiving

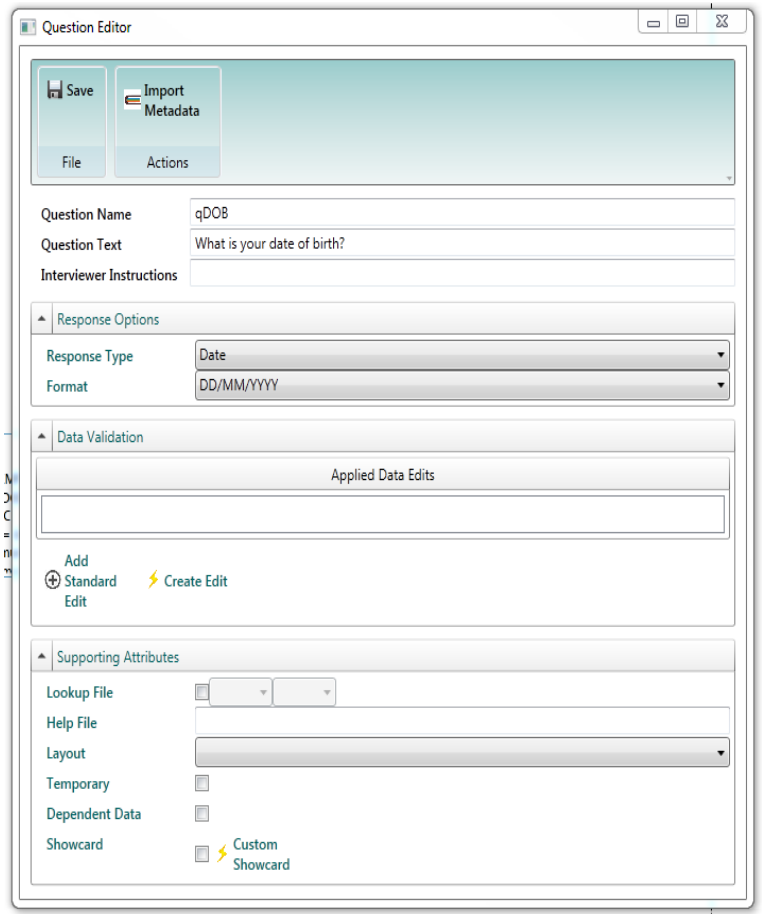
Key Result 3 – New tools

Questionnaire Editor



Key Result 3 – New tools

Questionnaire Editor



Facts and Figures (as of 2013)

1,008

Datasets

200

Series

20 - 40

Metadata Curators

219

Unique Portal Users

INSEE

- Represent questionnaires in DDI 3.2 XML
- Tested with ESA survey
  - the main French structural business survey
- Automatically generate:
  - Paper forms for respondents to complete
  - Blaise source code for data entry

INSEE – Next Steps

- ESA survey proved that DDI can be used for
  - Survey specification
  - Enabling survey generation
- Expanding use of DDI to create surveys in a project called Coltrane
- “DDI will be the standard of choice”

Statistics Denmark

- Managing Series and Study-level metadata
  - Using DDI Lifecycle
  - Describing concepts, populations
- Eurostat Quality reporting requirements
  - From DDI to ESMS, ESQRS SDMX formats
  - Automatic report generation

Statistics Denmark

- DDI Lifecycle is now used in their production environment
- Next steps include documenting
  - Variables and datasets
  - Classifications according to the Neuchatel model and DDI 3.2

Thank you

Jeremy Iverson
jeremy@colectica.com

Dan Smith
dan@colectica.com

	Web	colectica.com
	Blog	blogs.colectica.com
	Twitter	@Colectica
	YouTube	youtube.com/colectica