

PhUSE 2009

FROM RAW DATA TO SUBMISSION:  
A metadata driven process of data  
conversion to CDISC models

Presenter:

Dimitri Kutsenko

Entimo AG, Berlin, Germany

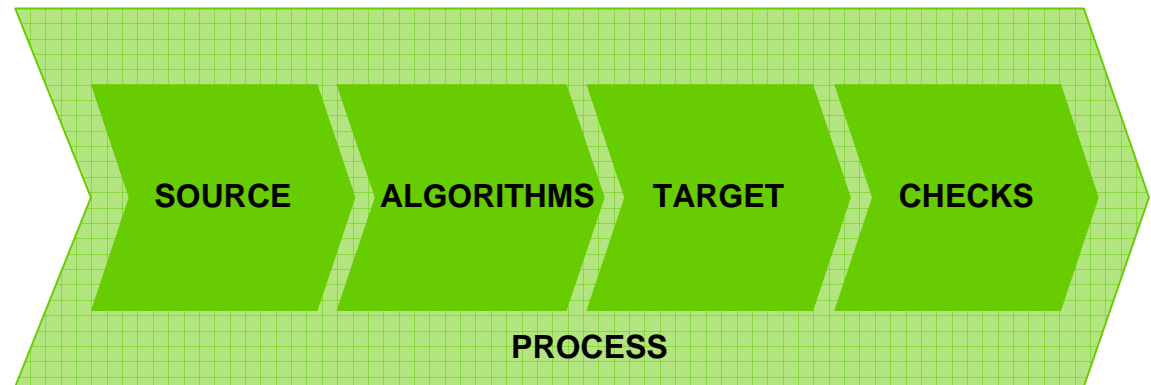
# From Raw Data to SDTM: Conservative Process



- **Inefficiencies:**
  - A lot of work done from the scratch
  - A lot of manual programming
  - “Copy&paste” reusability
  - Sequential process – unnecessary time losses
  - Many roles involved – communication frictions...
- **Smoother process?**

# Data Conversion Process: Generic View

- **Process:**



- **Metadata Challenges:**

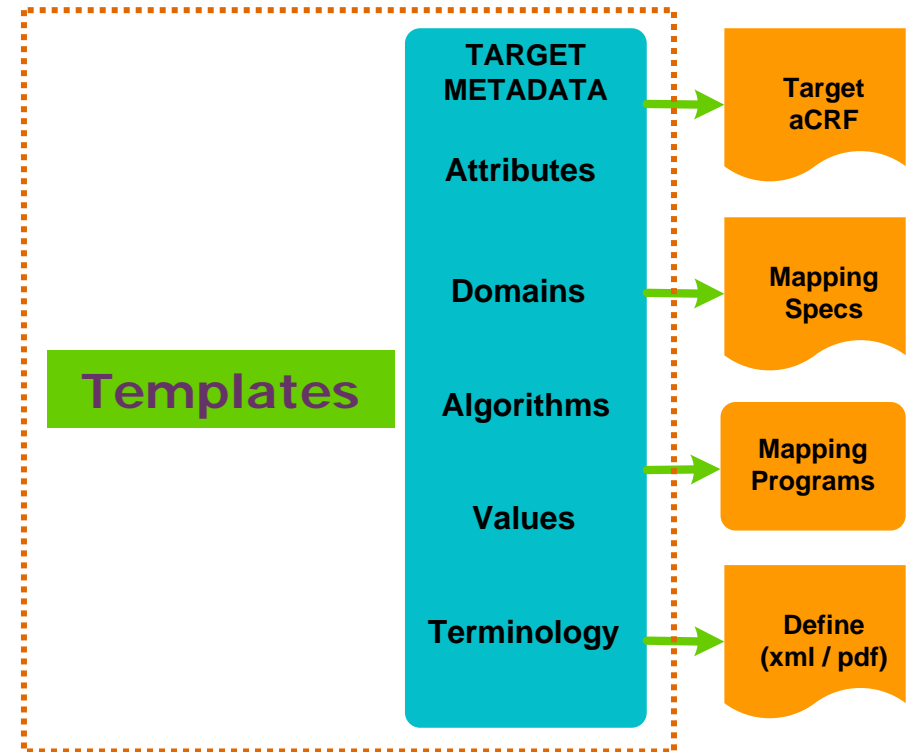
- Metadata variation vs. metadata “inflation”
- Model independence
- Interlinked, interdependent dimensions
- Automatic collection and processing
- Consistent management

# Source

- **Data:**
  - Delivered in different formats
  - Stored in different systems
  - “Frozen” – no edits required
  - Data’s metadata often not available
- **Requirements:**
  - Easy import procedure (files | database)
  - Data storage if required
  - Data access to avoid migration
  - Automatic metadata extraction

# Target

- **SDTM Domains:**
  - Publicly available
  - “Stable”
  - Adjustable (SDTM+)
  - Not suitable for automation
- **Requirements:**
  - Flexible definition
  - Processable format
  - Consistent management



**HIGH REUSABILITY  
POTENTIAL**

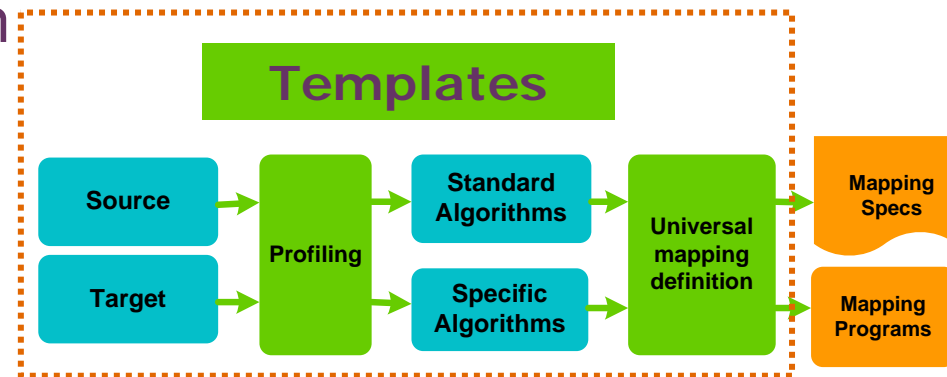
# Mapping Specifications And Programs

- **Generic Tasks:**

- Analyze sources and targets
- Develop mapping logic
- Create specification
- Develop mapping programs

- **Challenges:**

- Scarce experts
- Many roles involved
- Redundant information
- Repeated conversions



**HIGH OPTIMIZATION  
POTENTIAL**

# Data Conversion

- **Generic Task:**
  - Transform data
- **Challenges:**
  - Distributed sources
  - Many versions simultaneously
  - Manual process documentation inconsistent



**HIGH CENTRALIZATION POTENTIAL**

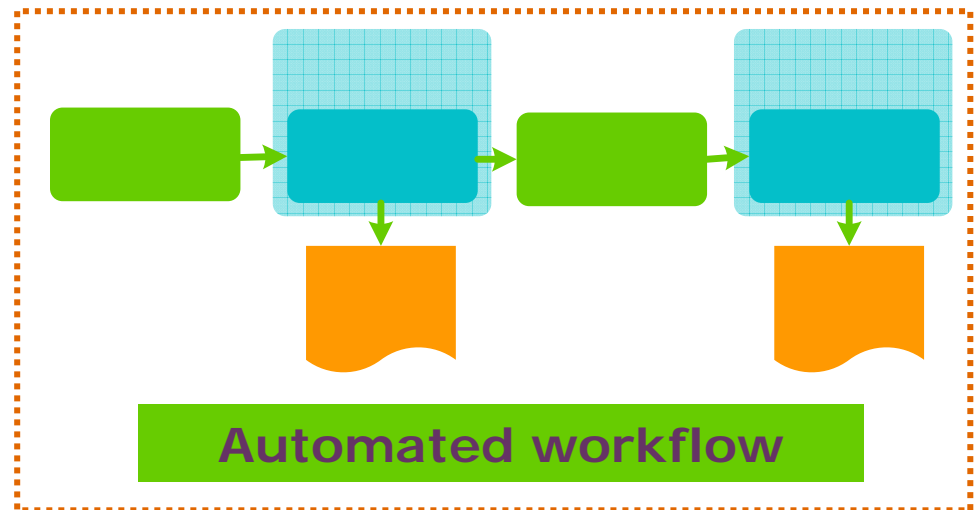
# Checks – Quality Control (QC)

- **Generic Tasks:**

- Review
- Source checks
- Content checks
- Formal checks

- **Challenges:**

- Manual interaction
- Late problem discovery
- Issue clarification

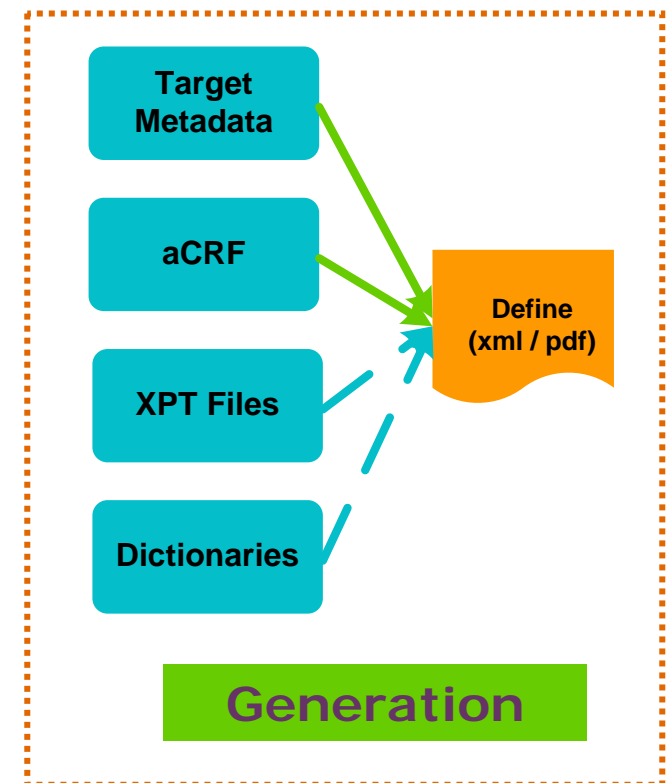


**HIGH  
AUTOMATION  
POTENTIAL**



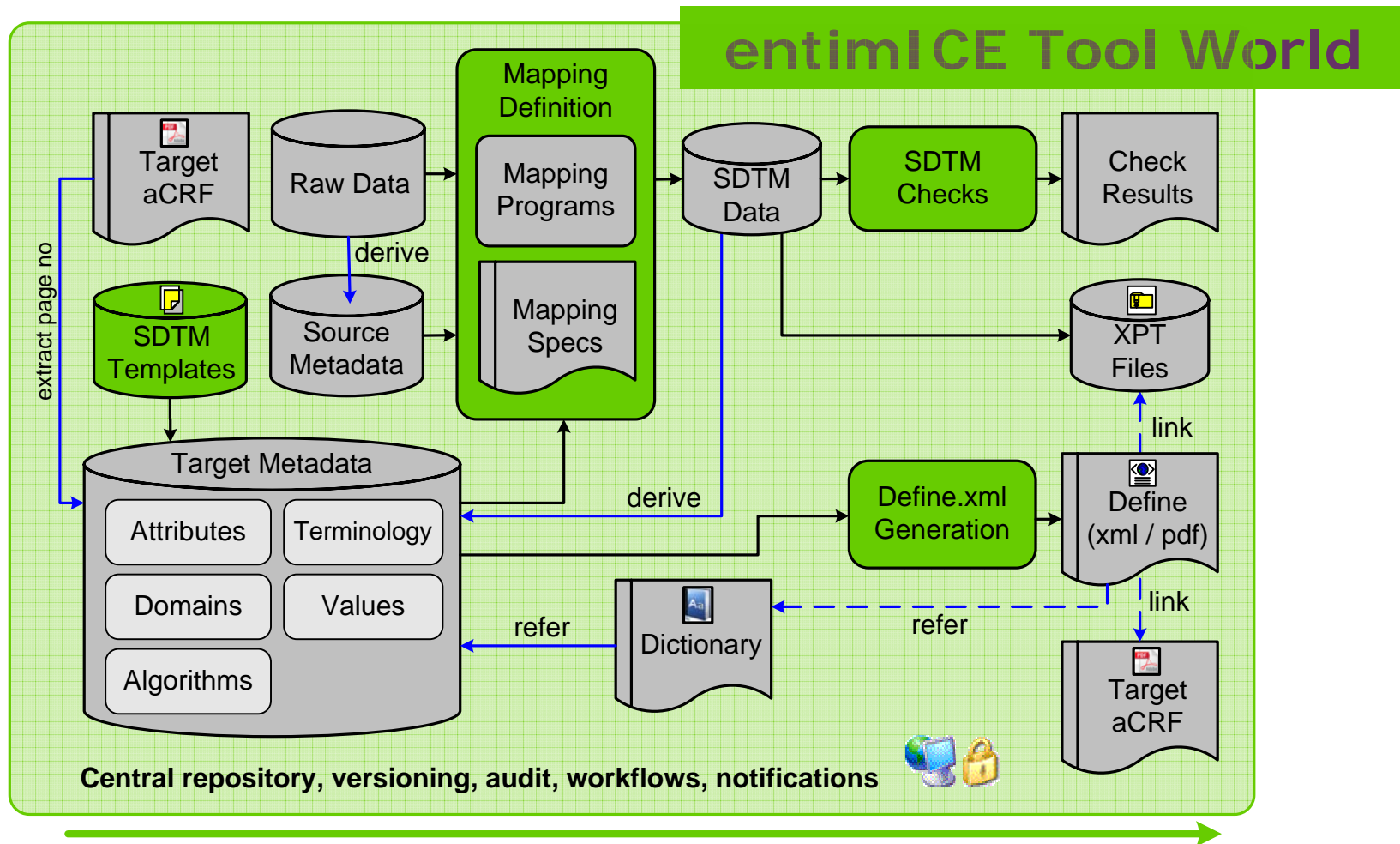
# Submission Overview (Define.xml)

- **Generic Task:**
  - Create define.xml
- **Challenges:**
  - Time consuming manual programming
  - Complex interlinking of information
  - Change-averse creation



**HIGH METADATA-USAGE POTENTIAL**

# From Raw Data To Submission: A Process Vision



# Introducing Global Standards

- Define company wide standards:
  - Supported models and versions
  - QC points and criteria
  - Workflows
- Provide for supported models:
  - Domain metadata templates
  - Standard algorithms
  - Standard mappings

END

Many thanks for your attention!  
Questions...?

**VISION STARTS NOW!**

Visit us at entimo's booth