



PhUSE 2012 DH07

Using Metadata to Integrate and Standardise Clinical Trial Data: The ICON Study Data Mapper

Existing situation

- Manual processes: High level of effort to specify, program and deliver data
- ETL custom written from scratch for each study
- Limited reuse of data deliveries within a study
- Limited reuse of programming across studies
- No single source of clinical data

Oracle ® Life Sciences Data Hub already in place

- Implemented in ICON Data Management in 2010

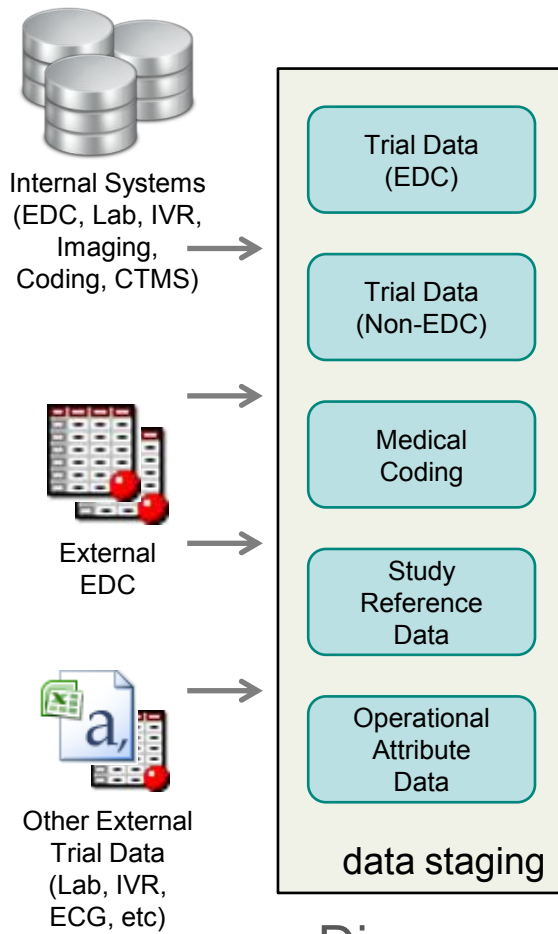
Project goals

- Build on the strengths of LSH (global repository of clinical data, program and data traceability)
- Replace multiple data flows with single, streamlined flow using standard data models
- Reengineer process for data specification and programming
- Facilitate reuse of code in other ICON departments (e.g., DM)

Three areas of focus

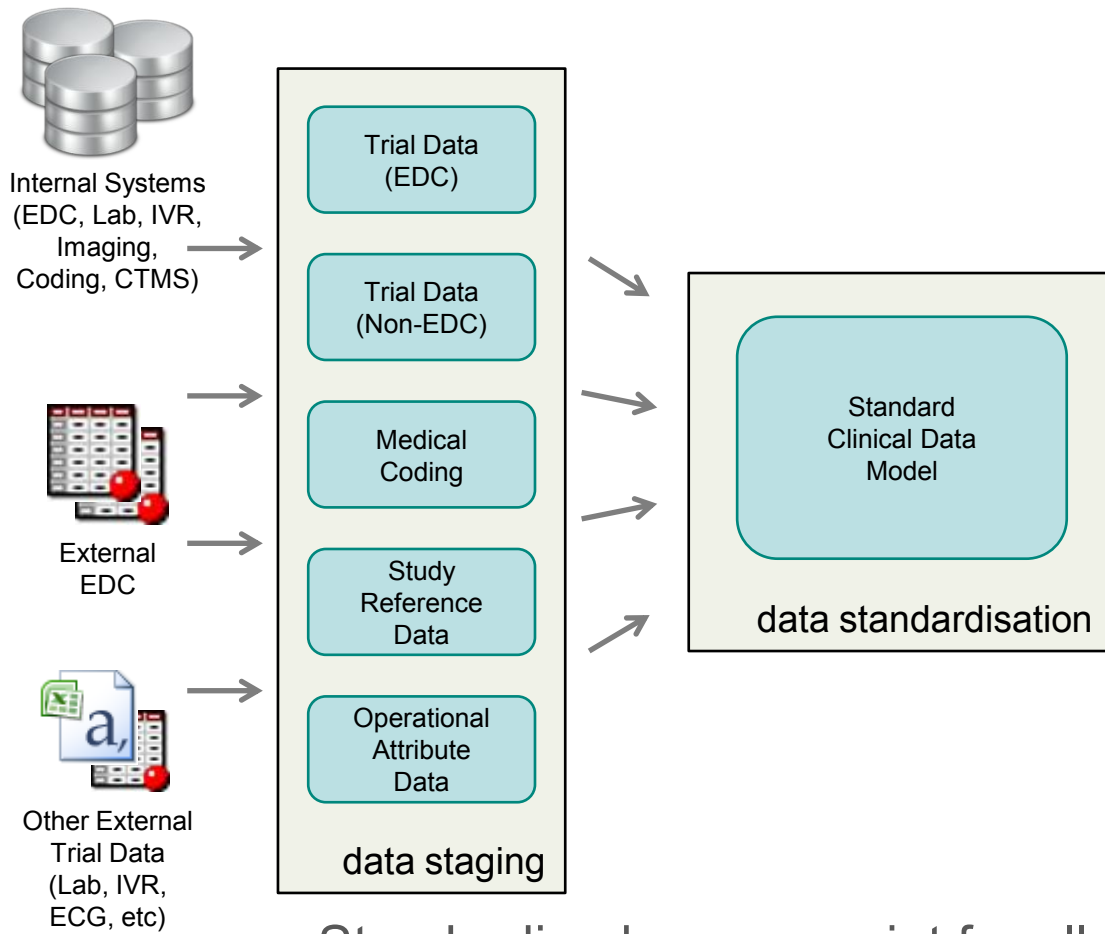
- Standard data flow
- Standard data models
- Standard tools

New Dataflow 1: Centralise



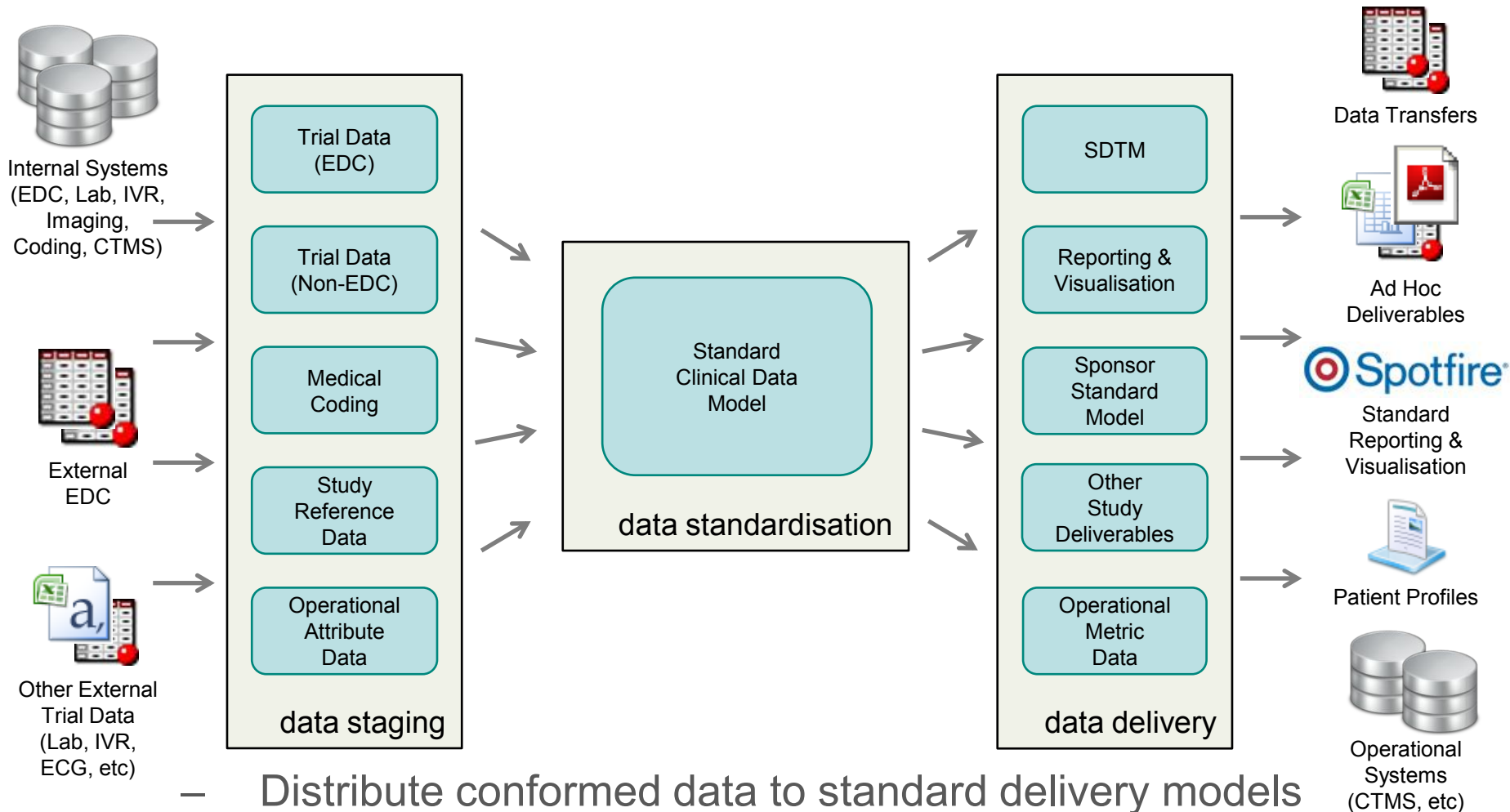
- Diverse sources with little standardisation
- Permanent staging area with full audit history of “raw” data

New Dataflow 2: Standardise



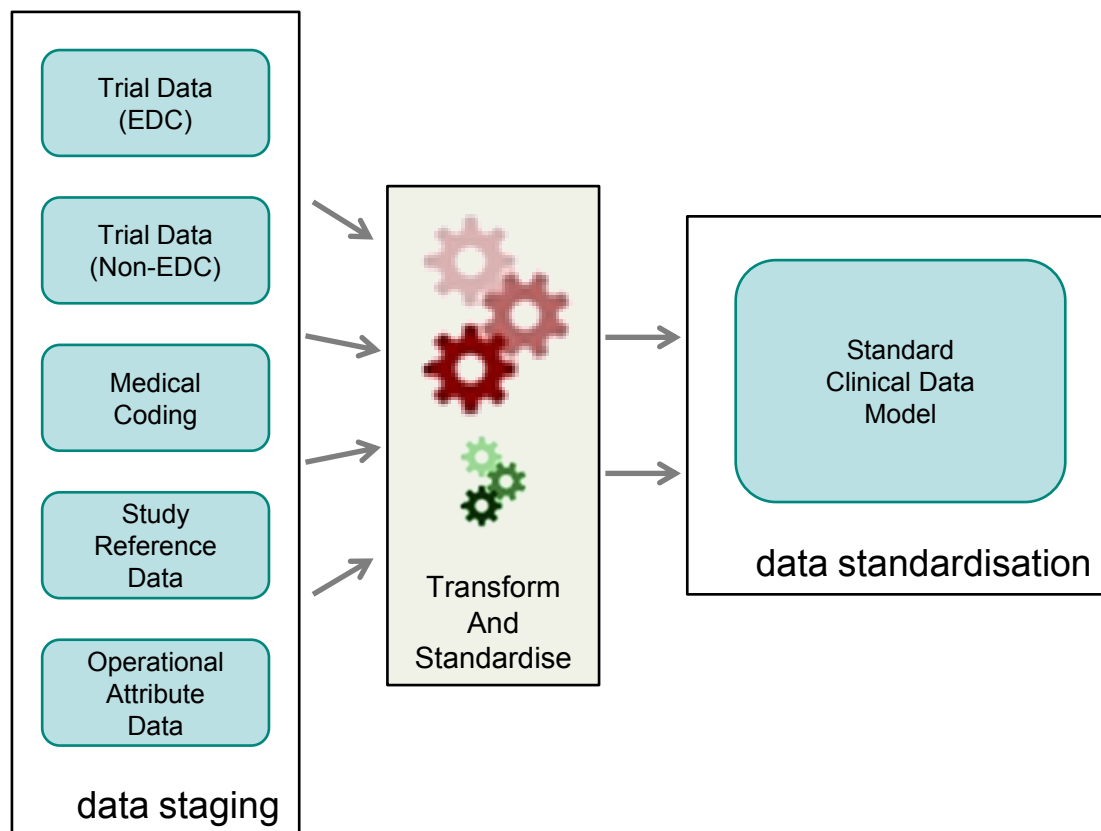
- Standardised access point for all study clinical data
- Standard clinical data model is an SDTM-like structure

New Dataflow 3: Deliver



- Distribute conformed data to standard delivery models
- Re-structured targeted copies of data for data consumers

Study Sources to Clinical Standard



- Involves a high level of study-specific ETL
- Limited use of standard code

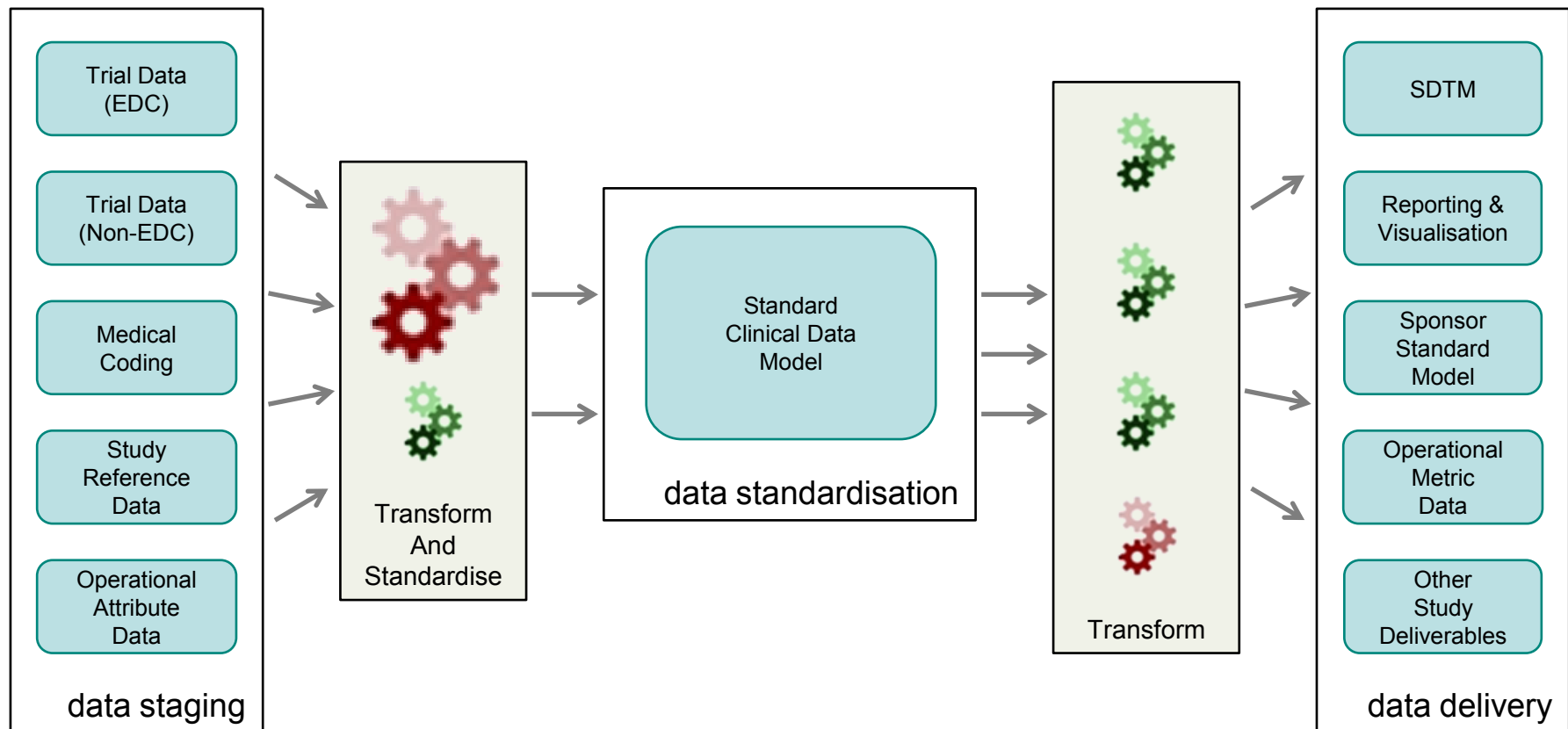


Standard Code



Custom Code

Clinical Standard to Delivery Standards



- Majority of ETL is standard library code
- Little or no study-specific ETL



Standard Code



Custom Code

Data standardisation work is done “up front”

- Most of the standardisation happens between source and the standard clinical model

Data standardisation must happen earlier in the life of study

- Users are reliant on the data being standardised to access new standard reports, etc.

Manual ETL development is costly and time consuming

Transform specifications are already required for traceability

- Describe related structures (source and target)
- Specify rules for transforming data between sources and targets

Use this asset as a source for metadata

Prerequisites for ETL derivation

- Complete, detailed specification of data structures
- Predictable approach to defining transformation rules
- Machine readable spec

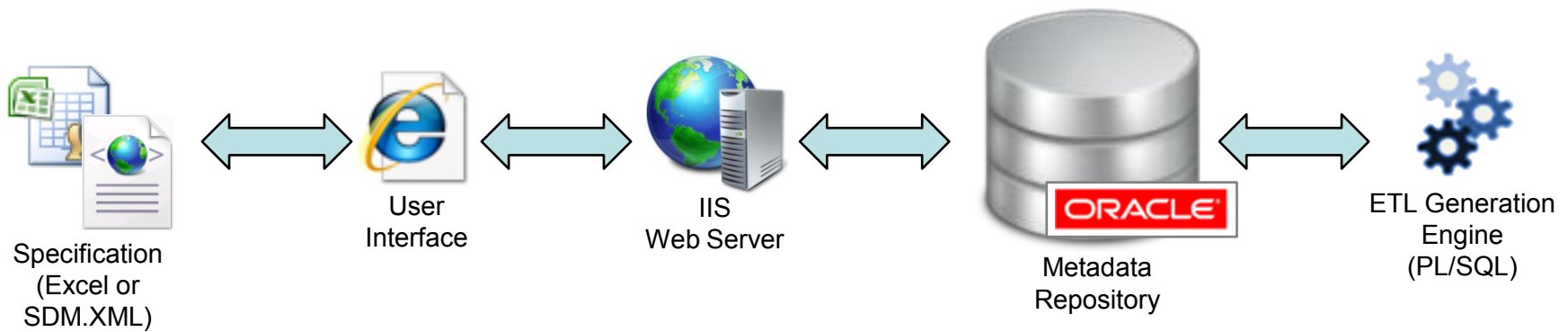
Study Data Mapper is a tool to

- Describe the metadata describing the tables, variables and value lists for sources and targets
- Specify how those source tables will be transformed into the target table structures
- Generate SAS code based on the recorded metadata

In-house software application

- Developed in partnership with BioPharm Systems
- Phase 1 delivered in March 2012, Phase 2 currently rolling out
- Development ongoing...

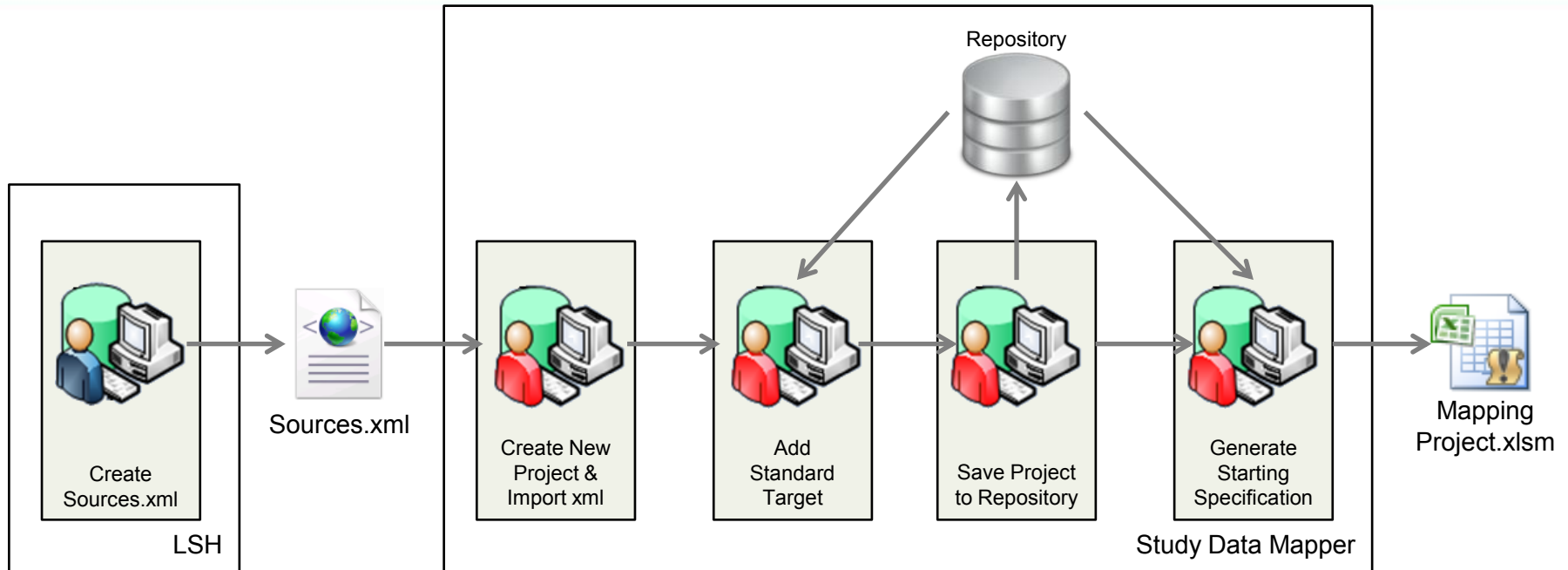
Number of alternative tools available



Components

- A web based user interface for managing mapping projects
- An integrated Excel spreadsheet for specifying metadata structures and transformation maps
- An Oracle database for securely storing the metadata repository
- An ETL generator that produces SAS PROC SQL code

Mapping Process: Project Setup



Project Setup

1. Create metadata for sources from LSH
2. Create mapping project and import sources metadata
3. Copy chosen target standard from repository
4. Save project to repository and generate starting specification

Project Metadata Walkthrough 1

Study and mapping project metadata

--Mapping Project	
Project Name :	A001_HUB
Project Version :	3
Description:	Map study A001 to Clinical standard model (HUB)

--Study Details	
Sponsor :	ACME
Program :	ABC
Study code :	A001
Protocol :	PRO_A
EDC System :	LSH
EDC System Version :	2.2
Default Tableset Name :	TARGET
EDC Study Location:	ACME>Oncology>A001>DATA>HUB

--Document Details (Replace with version number/date information)	
Authored by :	
Approved by :	
Current version :	
Date of completion :	
Date of approval :	

Screenshot from specification template

Table Metadata

Table set metadata

--SOURCE Tablesets			
Tableset Name	Author	Source System	Source Location
EDC	Paul Crean	RAVE	ACME>Oncology>A001>DATA>EDC Source
EXT	Paul Crean	LSH	ACME>Oncology>A001>DATA>External Source

--Target Tablesets			
Tableset Name	Author	Source System	Source Location
TARGET	Standards	Study Data Mapper	Standards Library
WORK	Mapper	Temp	Temp

Table metadata

--SOURCE Tables						
Order	TablesetName	Name	ShortName	IsRepeating	Label	Description
1	EDC	AE	AE	N	Adverse Events	Adverse Events
2	EXT	ADVEC	ADVEC	N	Coded Terms AE	Adverse Event Coded Terms

Target Tables					
Order	TablesetName	Name	ShortName	Label	Description
1	TARGET	STD_AE	AE	STD_AE	One record per adverse event per subject.
2	WORK	RFSTDTC	RFSTDTC	RFSTDTC	One record per subject

Variable metadata

--Source Variables							
Order	TableName	VariableName	ShortName	Pkey	Label	DataType	Length
1	EDC.AE	PROJECTID	projec00		projectid	Num	8
2	EDC.AE	PROJECT	project		project	Char	255
3	EDC.AE	STUDYID	studyid		Internal id for the study	Num	8
4	EDC.AE	ENVIRONMENTNAME	enviro00		Environment	Char	20
5	EDC.AE	SUBJECTID	subjec00		Internal id for the subject	Num	8
6	EDC.AE	STUDYSITEID	StudyS00		Internal id for study site	Num	8
7	EDC.AE	SUBJECT	Subject		Subject name or identifier	Char	50

Target Variables								
Order	TableName	VariableName	ShortName	Label	DataType	Length	Precision	
1	TARGET.STD_AE	STUDYID	STUDYID	Study Identifier	Char	15		
2	TARGET.STD_AE	DOMAIN	DOMAIN	Domain Abbreviation	Char	2		
3	TARGET.STD_AE	USUBJID	USUBJID	Unique Subject Identifier	Char	25		
6	TARGET.STD_AE	AESEQ	AESEQ	Sequence Number	Num	8		0
7	TARGET.STD_AE	AEGRPID	AEGRPID	Group ID	Char	40		
8	TARGET.STD_AE	AEREFID	AEREFID	Reference ID	Char	40		
9	TARGET.STD_AE	AESPID	AESPID	Sponsor-Defined Identifier	Char	40		
10	TARGET.STD_AE	AETERM	AETERM	Reported Term for the Adverse Event	Char	200		

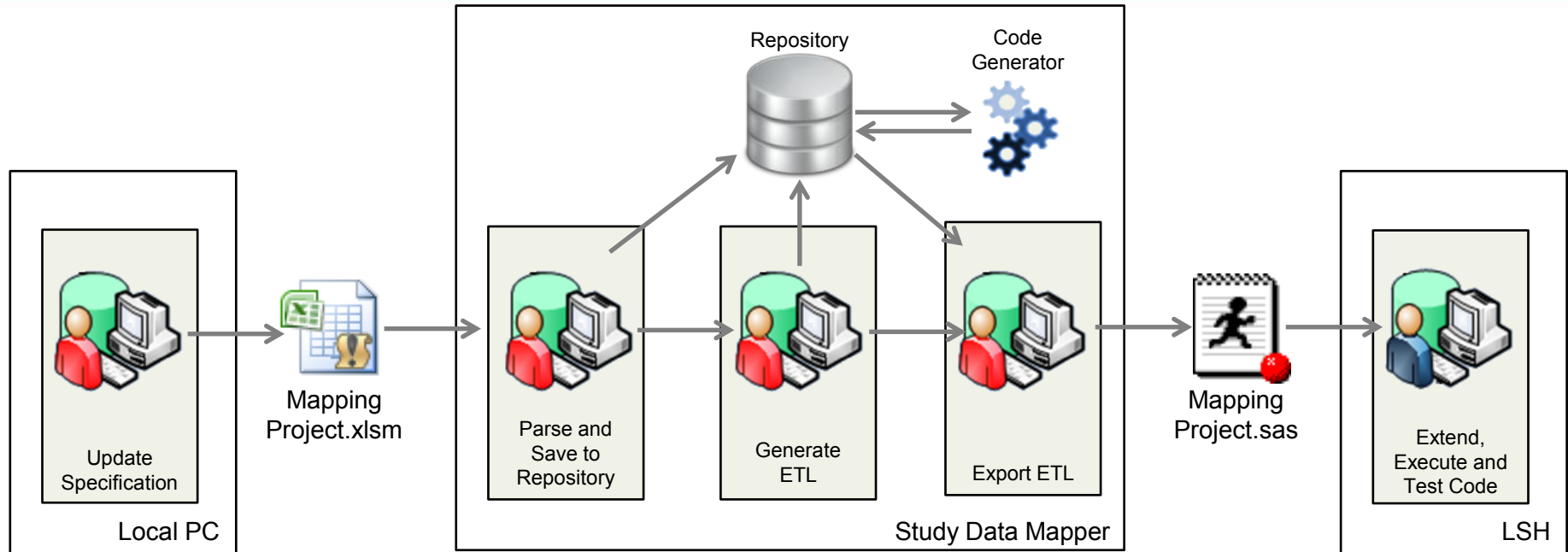
- Including SAS format, SDTM flag, supplemental qualifier flag

Value List metadata

--Source Value List Values			
SourceValueListName	Start Value	Code	Description
EDC.AESEV	Grade 1	1	
EDC.AESEV	Grade 2	2	
EDC.AESEV	Grade 3	3	
EDC.AESEV	Grade 4	4	
EDC.AESEV	Grade 5	5	

-Target Value List Values			
TargetValueListName	Start Value	Code	Description
TARGET.AESEV		1 MILD	Grade 1; 1
TARGET.AESEV		2 MODERATE	Grade 2; 2
TARGET.AESEV		3 SEVERE	Grade 3; 3

Mapping Process Phase 2: Project Mapping



Project Mapping

1. Update mapping specification
2. Parse and save specification to repository
3. Generate and export code
4. Execute and extend code with custom mapping operations

Table Mapping

- Supports horizontal and vertical mappings
- Temporary target tables and variables supported
- SQL JOIN completed by mapping author

Source to Target Mapping	
Target Table	Source Table
TARGET.STD_AE	EDC.AE
TARGET.STD_AE	TARGET.STD_DM
TARGET.STD_AE	EXT.ADVEC

--Conditions

Join : FROM EDC.AE AE
LEFT JOIN EXT.ADVEC ADVEC ON PUT(AE.RECORDID,8.) = ADVEC.EXTERNAL_ID
LEFT JOIN TARGET.STD_DM STD_DM ON AE.SUBJECT = STD_DM.SRC_SUBJID

Filter :

Variable Mapping

TARGET COLUMNS	Operation	Param1	Param2	Param3	Param4
TARGET.STD_AE.STUDYID	= ①	EDC.AE.PROJECT			
TARGET.STD_AE.USUBJID	CATX ②	' '	EDC.AE.PROJECT	EDC.AE.SITENUMBER	EDC.AE.SUBJECT
TARGET.STD_AE.AEBODSYS	%sdmUPCSTR ③	EXT.ADVEC.LEVEL_1_TEXT			
④ TARGET.STD_AE.AESTDTC	%sdmCREATEISODATE_RAVE	EDC.AE.AESTDAT			
TARGET.STD_AE.AEENDTC	%sdmCREATEISODATE_RAVE	EDC.AE.AEENDAT			
TARGET.STD_AE.AESTDY	%sdmDERIVESTUDYDY ④	TARGET.STD_AE.AESTDTC	TARGET.STD_DM.RFSTDTC		
TARGET.STD_AE.AEENDY	%sdmDERIVESTUDYDY	TARGET.STD_AE.AEENDTC	TARGET.STD_DM.RFSTDTC		
TARGET.STD_AE.AEDUR	%DeriveDurA0001 ⑤	TARGET.STD_AE.AEENDTC	TARGET.STD_AE.AESTDTC		

Variable mapping

- ① Simple assignments
- ② SAS functions
- ③ Global mapping macros
- ④ Target variables used as a source within the same table
- ⑤ Study specific mapping macros

PROC SQL

- Generated code creates and populates the target structures
- Each table map is generated separately and called in the order that takes care of dependencies

Portable code

- No dependency on Study Data Mapper, metadata or LSH
- Can be distributed to sponsors or regulatory authorities with a compiled macro catalog of mapping functions
- Clear, readable code: only basic PROC SQL knowledge required

Migration to LSH (including targets and ETL)

- Currently a manual process that takes about 20 minutes

Samples of Generated Code: Target Creation

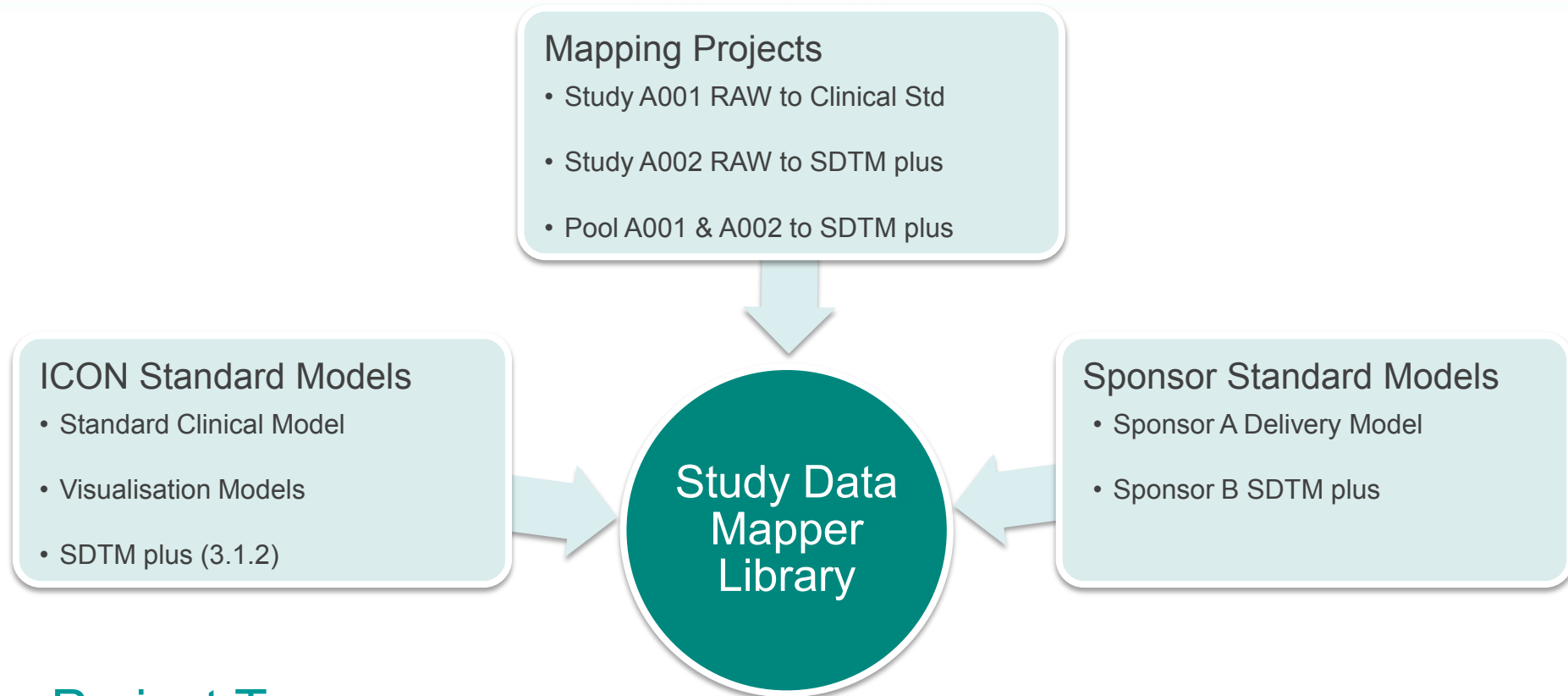
```
/*
  SAS Code Generated By Study Data Mapper

  Mapping Project: A001_HUB
  Map Set: A001_HUB version 1
  Generated By: ICON-EU\creanp
  Database: SDMDEV
  Date: 2012-10-10 17:48:28 +01:00
  Generation ID: 254294
*/

/* Processing table TARGET.STD_AE. */
%put NOTE: Processing TARGET.STD_AE. ;
%setLibname(tsname=EDC, sponsor=ACME, study=A001);
%setLibname(tsname=EXT, sponsor=ACME, study=A001);
%setLibname(tsname=TARGET, sponsor=ACME, study=A001);
PROC SQL;
CREATE TABLE TARGET.STD_AE(label="STD_AE"
, STUDYID VARCHAR(15) LABEL="Study Identifier"
, USUBJID VARCHAR(25) LABEL="Unique Subject Identifier"
, AEBODSYS VARCHAR(67) LABEL="Body System or Organ Class"
, AESTDTC VARCHAR(19) LABEL="Start Date/Time of Adverse Event"
, AEENDTC VARCHAR(19) LABEL="End Date/Time of Adverse Event"
, AESTDY NUMERIC(8,0) LABEL="Study Day of Start of Adverse Event"
, AEENDY NUMERIC(8,0) LABEL="Study Day of End of Adverse Event"
, AEDUR VARCHAR(12) LABEL="Duration of Adverse Event"
);
```

Samples of Generated Code: Target Population

```
INSERT INTO TARGET.STD_AE
(
  STUDYID
  ,USUBJID
  ,AEBODSYS
  ,AESTDTC
  ,AEENDTC
  ,AESTDY
  ,AEENDY
  ,AEDUR
)
SELECT
  AE.PROJECT AS STUDYID
  ,CATX('-', AE.PROJECT, AE.SITENUMBER, AE.SUBJECT) AS USUBJID
  ,%SDMUPCSTR(ADVEC.LEVEL_1_TEXT) AS AEBODSYS
  ,%SDMCREATEISODATE_RAVE(AE.AESTDAT) AS AESTDTC
  ,%SDMCREATEISODATE_RAVE(AE.AEENDAT) AS AEENDTC
  ,%SDMDERIVESTUDYDY(%SDMCREATEISODATE_RAVE(AE.AESTDAT), STD_DM.RFSTDTC) AS AESTDY
  ,%SDMDERIVESTUDYDY(%SDMCREATEISODATE_RAVE(AE.AEENDAT), STD_DM.RFSTDTC) AS AEENDY
  ,%DERIVEDURA0001(%SDMCREATEISODATE_RAVE(AE.AEENDAT), %SDMCREATEISODATE_RAVE(AE.AESTDAT)) AS AEDUR
FROM EDC.AE AE
LEFT JOIN EXT.ADVEC ADVEC
ON PUT(AE.RECORDID,8.) = ADVEC.EXTERNAL_ID
LEFT JOIN TARGET.STD_DM STD_DM
ON AE.SUBJECT = STD_DM.SRC_SUBJID
;
```



Project Types

- Data model projects (ICON standard, sponsor standard...)
- Mapping projects (standard-to-standard, raw-to-standard...)
- All projects are version controlled

Metadata is not a silver bullet

- The use of metadata has reduced programming effort on transforming studies by 90%
- BUT, it hasn't reduced the time it takes to transform a study

Why?

- We're standardising and delivering more data to a wider audience
- Programmers take 10-12 weeks to gain expertise in the new tools and processes
- A higher quality specification process takes longer
- Not enough reuse of metadata

More standards reuse required

- Metadata reuse depends on standards in the source systems

Contact Details

Paul Crean

Clinical Data Warehouse Mgr, ICONIK

ICON Clinical Research

Email: paul.crean@iconplc.com

Web: www.iconplc.com

Jennie McGuirk

Manager Clinical Data Delivery, ICONIK

Email: jennie.mcguirk@iconplc.com

