Tame the Traceability Monster

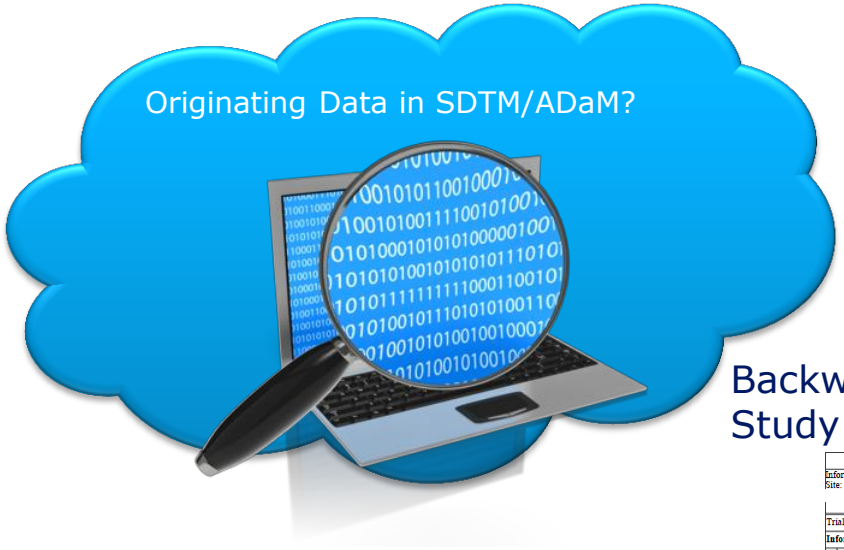
Mikkel Traun
Kirsten Walther Langendorf
PhUSE Single Day Event
28-May-2013

novo nordisk®

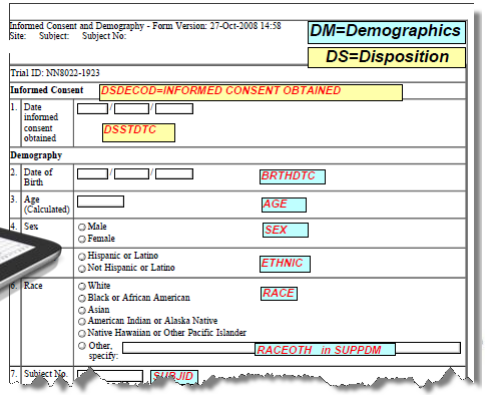
Outline

- Traceability from reviewer's and sponsor's perspectives
- CDISC project at Novo Nordisk – SDTM initial implementation
- Our approach to 'tame the traceability monster'

Traceability from reviewer's perspective

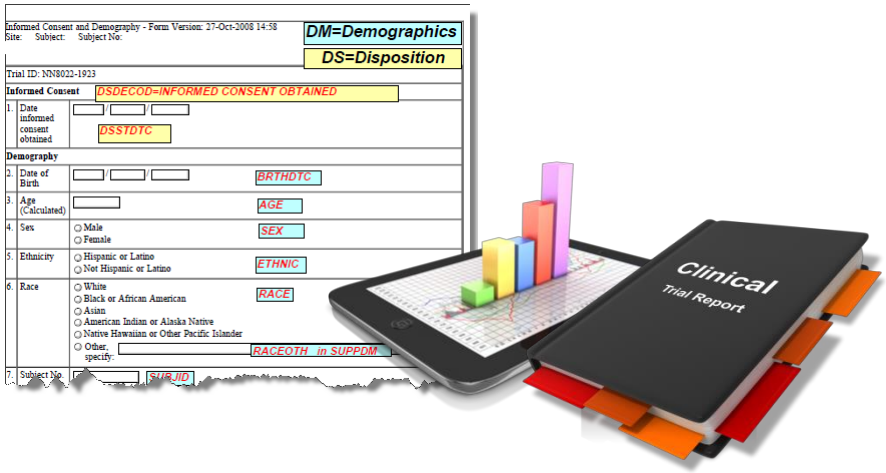


Backward Traceability:
Study Report – Analysis - eCRF



Traceability from sponsor's perspective

Forward Traceability:
eCRF – Analysis - Study Report



novo nordisk®

CDISC project at Novo Nordisk

2013 2014 2015 2016

2015 **CDISC terminology for all new trials**

2016 **CDISC SDTM/ADaM for all submissions**

Business Process Initiative

Future Business Process Initiatives/IT projects supporting CDISC

CDISC Controlled Terminology

Initial implementation
CDISC aligned CRFs on two selected project

Ramp-up Implementation
CDISC controlled terminology managed and used on all projects

Operations

CDISC SDTM

Initial Implementation
SDTM submission on one project (6 trials)
Test of various tools

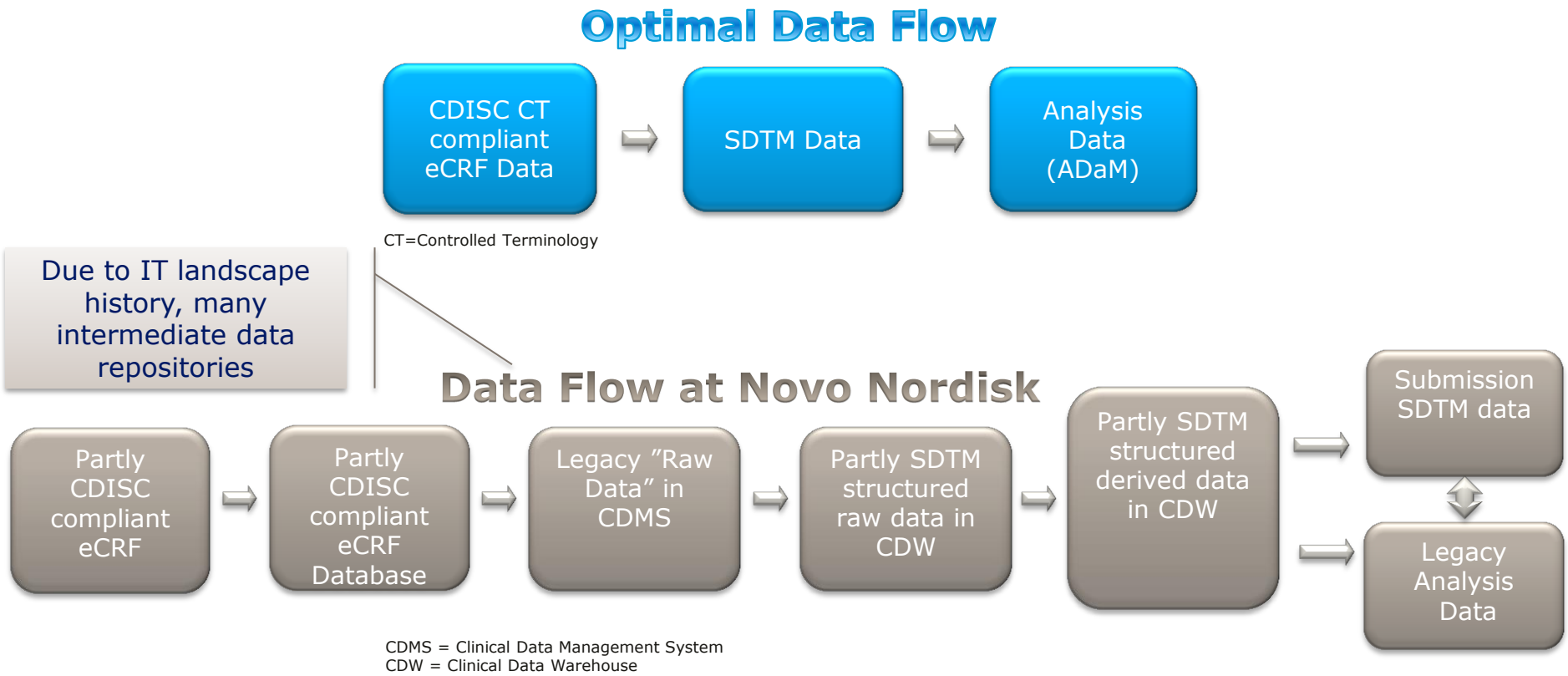
Ramp-up Implementation*
SDTM made for all projects
– approach for ADaM implementation

Operations

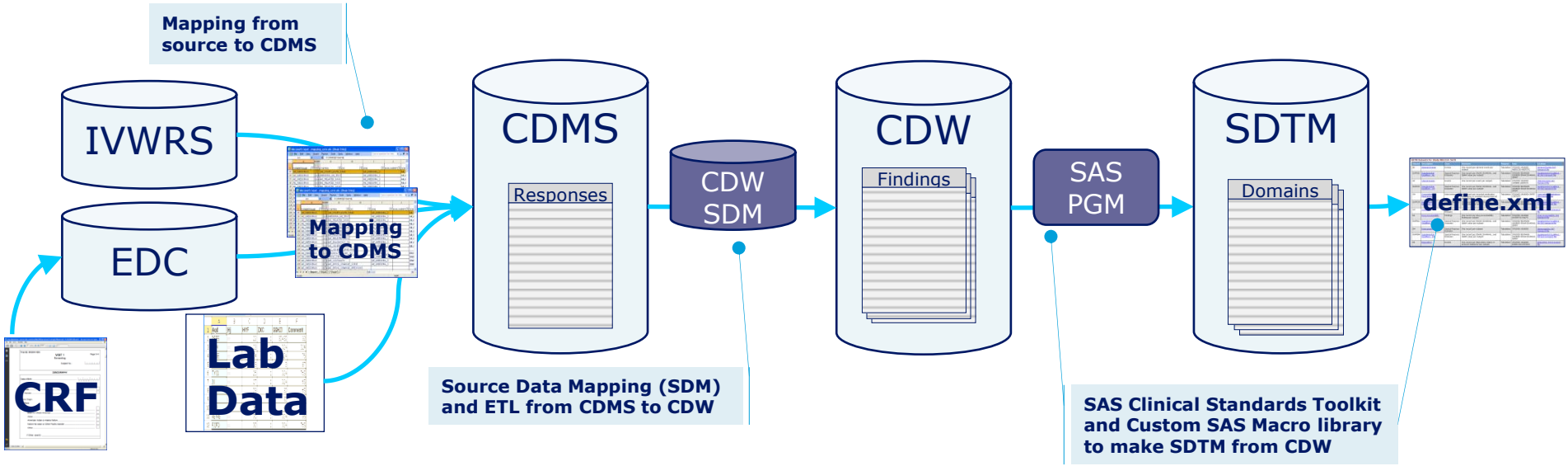
SDTM Solution based on SAS macro/programs and excel mapping sheets.
Tools: OpenCDISC Validator and SAS Clinical Standard Toolkit (CST)

* Analyse and decide on ADaM implementation

Traceability challenge



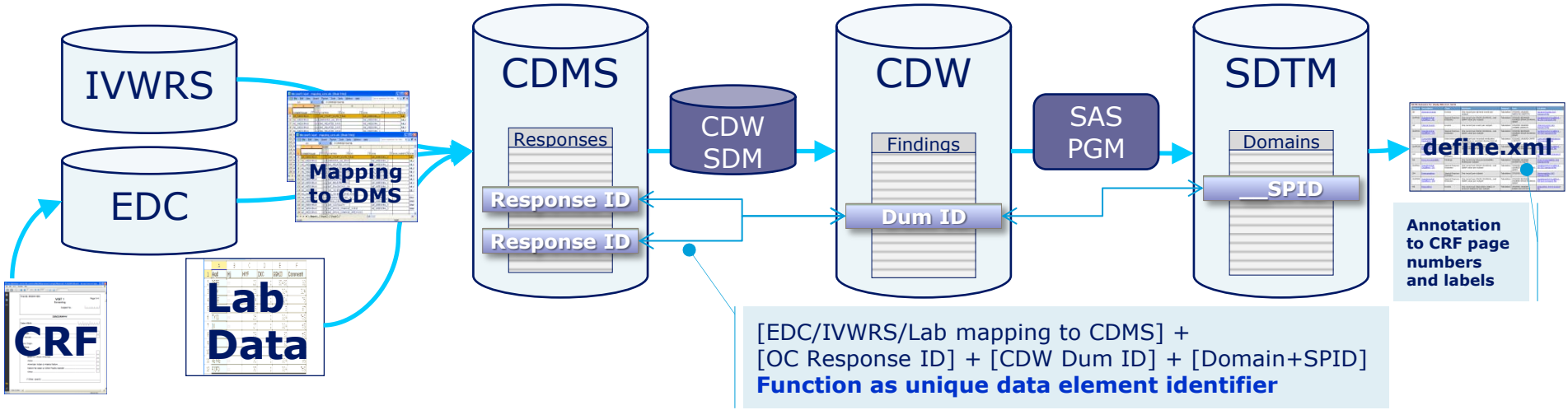
E2E data flow and intermediate data stores



- Physically in one normalised Question-Response table
  - Logically views by data collection forms
- Generalised into SDTM class datasets (Janus like)
- Submission ready data sets and supplemental qualifiers by domains

novo nordisk®

Traceability by unique data element identifiers

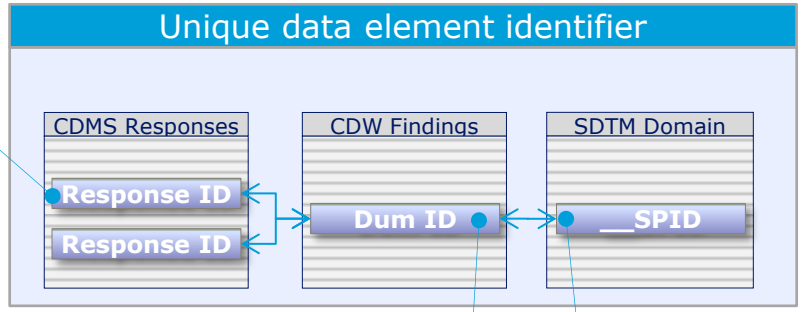


- Very difficult to document traceability based on different data standards applied in multiple systems
- Difficult to document traceability through multiple layers of mapping rules
- Simple and automated to document traceability by carrying **unique data element identifiers** through the data transformation

novo nordisk®

Unique data element identifier

Uniquely identifies data record in OC (CDMS), thereby trace to Data Collection Form and Question



SUBJ_TAB_TO_OC_RESPONSE_MAP

Filter and Sort | Query Builder | Data | Describe | Graph | Analyze

	response_id	dum_id	cdw_col_oid	sdm_qst_map_oid
1	80086493807	9467475565	322	4351878
2	80086494107	9467475565	321	4351895
3	80086494707	9467475565	319	4351874
4	80086494607	9467475565	468	4351899
5	103200506807	9467475566	318	4351880
6	103200507307	9467475566	317	4351883
7	103200506607	9467475566	315	4351885

Uniquely identifies data record in CDW, including:

- Reference to OC Response ID for a specific CDW column value
- Data snapshot
- Mapping rules

Uniquely identifies data record in SDTM domain, including:

- Reference to CDW Dum ID for a specific record, in a specific table, in specific snapshot
- Mapping rules, CDW to SDTM column

- **Data driven, record level, E2E traceability**

novo nordisk®

Define.xml – Backward traceability

SDTM Datasets for Study NN1234-5678

Dataset	Description	Class	Structure	Purpose	Keys	Location
AE	Adverse Events	Events	One record per adverse event per subject	Tabulation	STUDYID USUBJID AEDECOD AESTDTC	Adverse Events SAS transport file
SUPPAE	Supplemental Qualifiers - AE	Spe Dat				
CE	Clinical Events	Eve				
SUPPCE	Supplemental Qualifiers - CE	Spe Dat				
CM	Concomitant Medications	Int				
SUPPCM	Supplemental Qualifiers - CM	Spe Dat				
CO	Comments	Spe Dor				
DA	Drug Accountability	Fin				
SUPPDA	Supplemental Qualifiers - DA	Spe Dat				
DM	Demographics	Spe Dor				
SUPPDM	Supplemental Qualifiers - DM	Spe Dat				
DS	Disposition	Eve				

Demographics (DM) Demographics SAS transport file

Variable	Label	Key	Type	Length	Code List / Controlled Terms	Origin	Role	Source/Derivation/Comments
STUDYID	Study Identifier	1	text	40			Identifier	
DOMAIN	Domain Abbreviation		text	8			Identifier	
USUBJID	Unique Subject Identifier	2	text	40			Identifier	
BRTHDTC	Date/Time of Birth		text	64		eDT	RecordQualifier	IVWRS: Demography, Enter the subject's date of birth in 8 digit day, month, year
AGE	Age		float	8			RecordQualifier	
AGEU	Age Units		text	10	AGEU		VariableQualifier	
SEX	Sex		text	2	SEX	eDT	RecordQualifier	IVWRS: Demography, Enter the subjects gender. For male press 1, for female press
RACE	Race		text	80	RACE	CRF page 55	RecordQualifier	EDC: Informed Consent and Demography, Race
ETHNIC	Ethnicity		text	40	ETHNIC	CRF page 55	RecordQualifier	EDC: Informed Consent and Demography, Ethnicity
ARMCD	Planned Arm Code		text	20			RecordQualifier	
ARM	Description of Planned Arm		text	40			SynonymQualifier	
COUNTRY	Country		text	3	COUNTRY		RecordQualifier	
DMDTC	Date/Time of Collection		text	64			Timing	
DMDY	Study Day of Collection		integer	8			Timing	
DMSPID	Sponsor ID		text	40			Identifier	

Automatic reference to source system, thereby to Origin

Data driven reference to source as form and item annotation

Annotated CRF – Forward traceability

Informed Consent and Demography - Form Version: 27-Oct-2008 14:58
Site: Subject: Subject No:

DM=Demographics

DS=Disposition

Trial ID: NN1234-5678

Informed Consent — DSDECOD=INFORMED CONSENT OBTAINED

1.	Date informed consent obtained	___ / ___ / ___ DSSTDTC
Demography		
2.	Date of Birth	___ / ___ / ___ BRTHDTC
3.	Age (Calculated)	___ AGE
4.	Sex	○ Male ○ Female SEX
5.	Ethnicity	○ Hispanic or Latino ○ Not Hispanic or Latino ETHNIC
6.	Race	○ White ○ Black or African American ○ Asian ○ American Indian or Alaska Native ○ Native Hawaiian or Other Pacific Islander RACE ○ Other, specify: ___ RACEOTH in SUPPDM
7.	Subject No.	___ SUBJID

novo nordisk®

E2E Traceability excel report

E2E Traceability report listing distinct combinations of data elements from target systems to sources

E2E report can list data trace on various levels

eCRF to OC, CDW, SDTM dataset, column and Topic Code annotation
Trial ID : NN1234-5678 - Trial Definition ID : CTR - Trial Metadata Ver. : 4
Date : 13MAY13 16:04 - EDC Trial=Y - IVWRS Trial=Y

Src System	Form	Question	OC DCM Name	OC Question	CDW Dataset	CDW Column	CDW Column Label	SDTM Dataset	SDTM Column
EDC	Informed Consent and Demography	Age (Calculated)	DEMOGRAPHY	AGE_DERIVED	SUBJ	SRC_AGE_ON_CRF	SRC Age On CRF	DM	
EDC	Informed Consent and Demography	Ethnicity	DEMOGRAPHY	ETHNIC_DETAIL_CODE	SUBJ	SRC_ETHNICITY	SRC Ethnicity	DM	ETHNIC
EDC	Informed Consent and Demography	Race	DEMOGRAPHY	RACE_CODE	SUBJ	SRC_RACE	SRC Race	DM	RACE
EDC	Informed Consent and Demography	Race	DEMOGRAPHY	SPECIFY_OTHER	SUBJ	SRC_RACE_OTH	SRC Race Other	DM	
IVWRS	Demography	Enter the subject's date of birth in 8 digit day, month, year format	DEMOGRAPHY	BIRTH_DATE	SUBJ	SRC_DT_OF_BIRTH	SRC Date Of Birth	DM	BRTHDTC
IVWRS	Demography	Enter the subjects gender. For male press 1, for female press 2	DEMOGRAPHY	SEX_CODE	SUBJ	SRC_SEX	SRC Sex	DM	SEX

Src System	Form	OC DCM Name	CDW Dataset	SDTM Dataset
EDC	Informed Consent and Demography	DEMOGRAPHY	SUBJ	DM
EDC	Informed Consent and Demography	DEMOGRAPHY	SUBJ	DS
IVWRS	Demography	DEMOGRAPHY	SUBJ	DM
IVWRS	Demography	DEMOGRAPHY	SUBJ	DS

novo nordisk®

E2E Traceability

- Support populating Origin and Source comments in Define.xml
- Support QC of CRF annotation
- Support QC of generated SDTM data
- Support statistical programming

More tracing

- Clinical Trial Report (CTR) to Analysis data sets and SDTM
- Analysis data sets to SDTM
- Conversion from legacy codes to CDISC Controlled Terminology

novo nordisk®

Conclusion & Discussion (1)

- FDA: Use **CDISC standards** and controlled terminology E2E as the standards will provide **traceability**
- Internal data flow **not aligned** to CDISC standards nor consistently using controlled terminology makes this traceability a **challenge**
- Globally **unique object identifiers**, identifying each unique collected data element **support tracing the data** from it's origin in collection to target location in submission independent of complex mapping rules
- By extracting **distinct combinations** from source data elements to target data elements a semi-automated **data driven traceability** documentation can be produced

novo nordisk®

Conclusion & Discussion (2)

- Data driven E2E traceability could also have a **benefit** even with an optimal flow using **CDISC standards E2E**

- This is a **data driven E2E documentation**, but current structures for traceability documentation (Annotated CRF and Define.xml) do not fully utilise the potential E2E traceability
  - **Improvements** should be made to improve usability of E2E tracing for data providers and reviewers
- E2E traceability should also be used to **optimise internal clinical standards management** as well as the **clinical data process**

novo nordisk®


THANK
YOU