



**2023**  
**EUROPE**  
**INTERCHANGE**  
**COPENHAGEN | 26-27 APRIL**



## **Tales from the journey of a CORE developer**

**Jozef Aerts**  
**XML4Pharma**



## Meet the Speaker

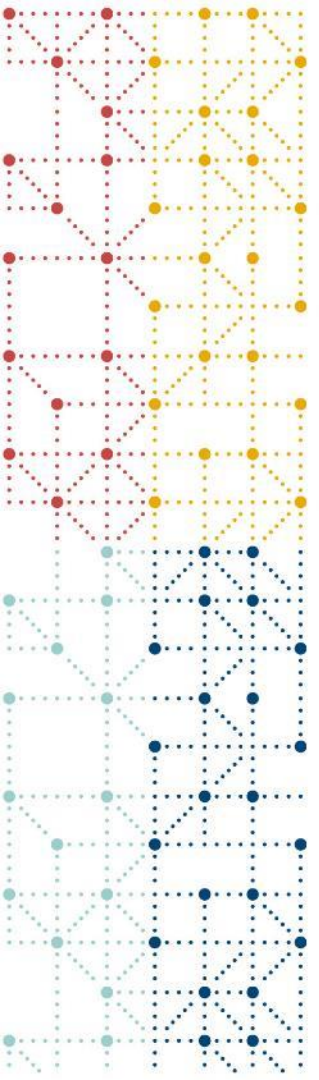
Jozef Aerts

**Organization: XML4Pharma**

Jozef is CEO of a small consultancy and software company specialized in CDISC standards. He is also a CDISC volunteer for more than 20 years and CDISC instructor.

# Disclaimer and Disclosures

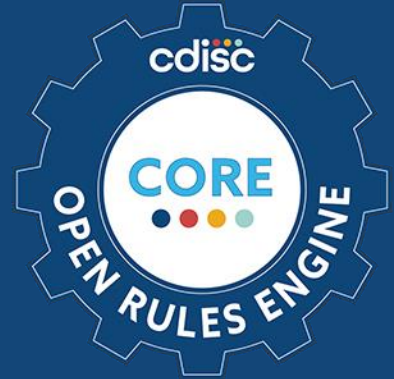
- *The views and opinions expressed in this presentation are those of the author(s) and do not necessarily reflect the official policy or position of CDISC.  
(but some of them should ...)*
- *{Please disclose any financial relationship or conflict of interest relevant to this presentation here OR}*
- *The author(s) have no real or apparent conflicts of interest to report.*



## Agenda

1. Jozef and CORE: How it started
2. CORE implementation in Software
3. Conclusions

# Jozef and CORE: How it started ...





## Some years ago ...

# We developed the "Open Rules for CDISC Standards" (ORCS)

Executable rules were developed using XQuery and made publicly available (as Open Source)

### Disadvantages:

- XQuery only works on XML
- FDA however has put the introduction of CDISC's Dataset-XML as a transport format on ice.
- Argumentation: file sizes
- Real reason: fear for change



#### The "Open Rules for CDISC Standards" initiative

**UPDATE!** [FDA SEND 3.0/3.1 validation rules 2019](#) (last update: 2020-07-01)  
**UPDATE!** [CDISC SDTM validation rules 2020 \(v.1.1\)](#) (last update: 2020-08-23)  
**UPDATE!** [PMDA SDTM validation rules 2019 \(v.2.0\)](#) (last update: 2020-08-22)  
**UPDATE!** [FDA SDTM validation rules 2018](#) (last update: 2020-08-22)

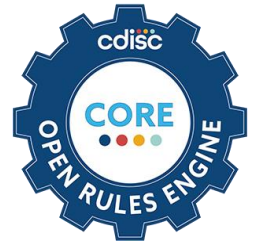
These rule sets have been updated to work with [define.xml 2.1](#).

# 6 years later ... CDISC CORE

## How can I contribute?

- I am pretty bad in Python coding
- Many volunteers already for SDTM
- ADaM is not my thing
- Define-XML will come later

**So I volunteered to implement the SENDIG (3.1) rules!**







# SENDIG-3.1 Rules

SEND\_Conformance\_Rules\_v4.0.xlsx \* - PlanMaker

CDISC SEND Rule ID

A	B	C	D	E	F	G	H	I	J	K	L
CDISC SEND Rule ID	Class	Domain	Variable	Plain Text Explanation of Condition/Rule	Condition	Rule	Cited Document	Cited Document Section	Cited Item	Cited Guidance	SENDIG Version
1	ALL	ALL	NA	SEND datasets are named to be consistent with the domain code (xpt file name is the domain abbreviation); exceptions to this rule are RELREC, SUPP, and POOLDEF.	not SUPP OR RELREC OR POOLDEF	SEND dataset file name is the domain abbreviation.	SENDIG v3.0	4.1.4	Text	SDTM datasets are normally named to be consistent with the domain code.	SENDIG v3.0
2	ALL	ALL	NA	SEND datasets are named to be consistent with the domain code (xpt file name is the domain abbreviation); exceptions to this rule are RELREC, SUPP, and POOLDEF.	not SUPP OR RELREC OR POOLDEF	SEND dataset file name is the domain abbreviation.	SENDIG v3.1	4.1.4	Text	SDTM datasets are normally named to be consistent with the domain code.	SENDIG v3.1
3	ALL	ALL	NA	SEND datasets are named to be consistent with the domain code (xpt file name is the domain abbreviation); exceptions to this rule are RELREC, SUPP, and POOLDEF.	not SUPP OR RELREC OR POOLDEF	SEND dataset file name is the domain abbreviation.	SENDIG v3.1.1	4.1.4	Text	SDTM datasets are normally named to be consistent with the domain code.	SENDIG v3.1.1
4	ALL	ALL	NA	SEND datasets are named to be consistent with the domain code (xpt file name is the domain abbreviation); exceptions to this rule are RELREC, SUPP, and POOLDEF.	not SUPP OR RELREC OR POOLDEF	SEND dataset file name is the domain abbreviation.	SENDIG v3.1	4.1.4	Text	SDTM datasets are normally named to be consistent with the domain code.	SENDIG-DART v1.1
5	ALL	ALL	GEN	Variable Name limited to 8 characters		Variable name length <= 8	SENDIG v3.0	3.2.2	Text	The Variable Name (limited to 8 characters for compatibility with the SAS v5 Transport format).	SENDIG v3.0
6	ALL	ALL	GEN	Variable Name limited to 8 characters		Variable name length <= 8	SENDIG v3.1	3.2.2	Text	The Variable Name (limited to 8 characters for compatibility with the SAS v5 Transport	SENDIG v3.1

<https://www.cdisc.org/standards/foundational/send/conformance-rules-v2-0-sendig-v3-0-and-sendig-v3-1>



# Development and Testing using the Rule Editor

```
EDIT TEST
1 Check:
2 any:
3   - all:
4     - name: ARMCD
5       operator: non_empty
6     - name: ARMCD
7       operator: is_not_contained_by
8         value: $ta_armcd
9   - all:
10    - name: ARM
11      operator: non_empty
12    - name: ARM
13      operator: is_not_contained_by
14        value: $ta_arm
15 Operations:
16   - domain: TA
17     id: $ta_armcd
18     name: ARMCD
19     operator: distinct
20   - domain: TA
21     id: $ta_arm
22     name: ARM
23     operator: distinct
24 Core:
25   Id: CDISC.SENDIG.224
26   Version: '1'
27   Status: Draft
28 Description: When study does not use multi-stage arm assignments and ARMCD must be present in TA.ARMCD. This rule has been executed to identify when ARMCD is not present in TA.ARMCD and therefore acknowledges that may be recorded when multi-stage arm assignments are in use.
29 Outcome:
30   Message: ARMCD is not present in TA.ARMCD
31 Rule Type: Record Data
32 Sensitivity: Record
33 Authorities:
34   - Organization: CDISC
```

EDIT TEST

---

- ✓ Validate YAML Syntax
- ✗ Validate YAML against Schema
- ✓ Convert YAML to JSON Rule

---

- ✓ Load Test Data

TEST DATASETS FILE...

Filename: unit-test-coreid-SENDIG-224-negative.xlsx  
Filetype: application/vnd.openxmlformats-officedocument.spreadsheetml.sheet  
Size in bytes: 13049  
Last modified date: 9/27/2022, 2:20:10 PM

```
[ 2 items
  0: { 5 items
    "filename": "dm_oggs"
    "label": "Demographics"
    "domain": "DM"
    "variables": [ 26 items
      0: { 4 items
        "name": "STUDYID"
        "label": "Study Identifier"
        "type": "Char"
        "length": 50
      }
      1: { 4 items
        "name": "DOMAIN"
        "label": "Domain Abbreviation"
        "type": "Char"
        "length": 50
      }
      2: { 4 items
        "name": "USUBJID"
        "label": "Unique Subject Identifier"
        "type": "Char"
        "length": 50
      }
      3: { 4 items
        "name": "SUBJID"
        "label": "Subject Identifier for the Study"
        "type": "Char"
        "length": 50
      }
      4: { 4 items
        "name": "RFSIDTC"
        "label": "Subject Reference Start Date/Time"
        "type": "Char"
        "length": 50
      }
    ]
  }
  1: { 4 items
    "name": "ARMCD"
    "label": "Arm CD"
    "type": "Char"
    "length": 50
  }
]
```

Results Negatives 3 Scope Skip 1

# Managing the rules in JIRA

All content on this Wiki is non-binding and any individual opinions expressed should not be considered indicative of the policies or positions of CDISC or any other organization.

CORE Rules / CORERULES-661  
SEND225

Edit Add comment Assign More Undo QC Start QC

## Details

Type:  Review Comments Status: **AWAITING QC** (View Workflow)  
Priority:  To be assigned Resolution: Unresolved  
Affects Version/s: None Fix Version/s: None  
Component/s: SENDIG v3.1  
Labels: None  
Executability: Fully Executable

## Description

[Click to add description](#)

## Attachments

Drop files to attach, or browse.

## Activity

All Comments History Activity

Jozef Aerts added a comment - 01/Sep/22 1:18 PM

Rule = "If present, the ARM value in DM for a given subject must exist in TA."

P.S. Are there exceptions? Like "NOT ASSIGNED"?

## People

Assignee: Jozef Aerts  
Reporter: Amy Palmer  
Votes: 0 [Vote for this issue](#)  
Watchers: 2 [Stop watching this issue](#)

## Dates

Created: 23/Aug/22 3:55 PM  
Updated: 22/Mar/23 7:48 PM

## Agile

[View on Board](#)

# My conclusions so far regarding Rules

- When rules were developed in the past, we did not think about computability
- Some rules express expectations (or best practices) rather than rules

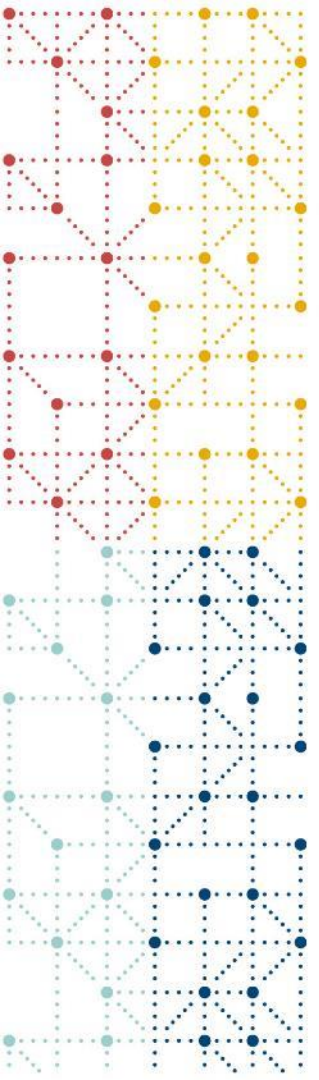
All Comments History Activity

▼  Jozef Aerts added a comment - 01/Sep/22 1:28 PM

Rule = "The values of EPOCH should provide a description of a time period that is independent of the value of ARM."

This rule is not executable as it requires semantic analysis.

- Sometimes there is "wobble room"
- We need to get better in defining exact rules (pre-condition / post-condition)
- Implementation: avoid "*false positives*"!  
(under-reporting rather than over-reporting)

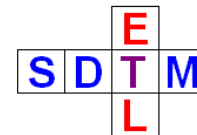


# CORE Implementation in Software: SDTM-ETL

It's easy if you know how ...

# CORE Implementation in Software

## What is SDTM-ETL ?



SDTM-ETL - version 4.1

File Edit View Navigate Insert Transform Validate CDISC Library Options About

ODM

- Study
  - GlobalVariables
  - BasicDefinitions
  - MetaDataVersion : Version 1.1.0
    - Protocol
      - StudyEventDef: Pre-treatment
      - StudyEventDef: Post-treatment
        - FormDef: Visit Form
        - FormDef: Adverse Events
          - ItemGroupDef: Common
          - ItemGroupDef: Adverse Events
            - ItemDef: Therapeutic Area
            - ItemDef: Protocol Number
            - ItemDef: Country
            - ItemDef: Record status, 5 levels, internal
            - ItemDef: (AE) Line Number
            - ItemDef: Conmed Indication
            - ItemDef: Start Month - Enter Two Digits 01-12
            - ItemDef: Start Day - Enter Two Digits 01-31
            - ItemDef: Start Year - Enter Four Digit Year
            - ItemDef: Derived Start Date
            - ItemDef: Stop Month - Enter Two Digits 01-12
            - ItemDef: Stop Day - Enter Two Digits 01-31
            - ItemDef: Stop Year - Enter Four Digit Year
            - ItemDef: Derived Stop Date
            - ItemDef: Severity
            - ItemDef: Relationship to study drug
            - ItemDef: Outcome
            - ItemDef: Actions taken re study drug
            - ItemDef: Actions taken, other
          - FormDef: Concom Meds

Domains (ItemGroups)

Variable	Variable	Variable	Variable	Variable	Variable	Variable	Variable	Variable
UID	DM.RFSTDTCT	DM.RFENDTCT	DM.SITEID	DM.INVID	DM.INVNAM	DM.BRTHDTC	DM.AGE	DM.
SGRPID	TS.TSPARMCD	TS.TSPARM	TS.TSVAL					
SESEQ	SE.ETCD	SE.ELEMENT	SE.SESTDTCT	SE.SEENDTCT	SE.TAETORD	SE.EPOCH	SE.SEUPDES	
VSITNUM	SV.VISIT	SV.VISITDY	SV.SVSTDTCT	SV.SVENDTCT	SV.SVSTDY	SV.SVENDY	SV.SVUPDES	
EXSEQ	EX.EXGRPID	EX.EXSPID	EX.EXTRT	EX.EXCAT	EX.EXSCAT	EX.EXDOSE	EX.EXDOSTXT	EX.
CMSEQ	CM.CMGRPID	CM.CMSPID	CM.CMTRT	CM.CMMODIFY	CM.CMDECOD	CM.CMCAT	CM.CMSCAT	CM.
SUSEQ	SU.SUGRPID	SU.SUSPID	SU.SUTRT	SU.SUMODIFY	SU.SUDECOD	SU.SUCAT	SU.SUSCAT	SU.
ESEQ	AE.AEGRPID	AE.AEREFID	AE.AESPID	AE.AETERM	AE.AEMODIFY	AE.AEDECOD	AE.AECAT	AE.
DSSEQ	DS.DSGRPID	DS.DSREFID	DS.DSSPID	DS.DSTERM	DS.DSDECOD	DS.DSCAT	DS.DSSCAT	DS.
DVSEQ	DV.DVREFID	DV.DVSPID	DV.DVTERM	DV.DVDECOD	DV.DVCAT	DV.DVSCAT	DV.EPOCH	DV.
CESEQ	CE.CEGRPID	CE.CEREFID	CE.CESPID	CE.CETERM	CE.CEDECOD	CE.CECAT	CE.CESCAT	CE.
MHSEQ	MH.MHGRPID	MH.MHREFID	MH.MHSPID	MH.MHTERM	MH.MHMODIFY	MH.MHDECOD	MH.MHCAT	MH.
EGSEQ	EG.EGGRPID	EG.EGREFID	EG.EGSPID	EG.EGTERM	EG.EGDECOD	EG.EGCAT	EG.EGSCAT	EG.
IESEQ	IE.IESPID	IE.IETESTCD	IE.IETEST	IE.IECAT	IE.IEDECOD	IE.IEORRES	IE.IESTRESC	IE.V
LBSEQ	LB.LBGRPID	LB.LBREFID	LB.LBSPID	LB.LBTERM	LB.LBDECOD	LB.LBCAT	LB.LBSCAT	LB.I
PESEQ	PE.PEGRPID	PE.PESPID	PE.PETESTCD	PE.PETEST	PE.PEMODIFY	PE.PECAT	PE.PESCAT	PE.
SCSEQ	SC.SCGRPID	SC.SCSPID	SC.SCTESTCD	SC.SCTEST	SC.SCCAT	SC.SCSCAT	SC.SCORRES	SC.
VSEQ	VS.VSGRPID	VS.VSSPID	VS.VSTESTCD	VS.VSTEST	VS.VSCAT	VS.VSSCAT	VS.VSPOS	VS.V
IBIID	CO.COSEQ	CO.IDVAR	CO.IDVARVAL	CO.COREF	CO.COVAL	CO.COVAL	CO.CODTC	
QSEQ	QS.QSGRPID	QS.QSSPID	QS.QSTESTCD	QS.QSTEST	QS.QSCAT	QS.QSSCAT	QS.QSORRES	QS.
DASEQ	DA.DAGRPID	DA.DAREFID	DA.DASPID	DA.DATESTCD	DA.DATEST	DA.DACAT	DA.DASCAT	DA.
PCSEQ	PC.PCGRPID	PC.PCREFID	PC.PCSPID	PC.PCTESTCD	PC.PCTEST	PC.PCCAT	PC.PCSCAT	PC.
PPSEQ	PP.PPGRPID	PP.PPREFID	PP.PPTESTCD	PP.PPTEST	PP.PPCAT	PP.PPORRES	PP.PPORRESU	PP.
MBSEQ	MB.MBGRPID	MB.MBREFID	MB.MBSPID	MB.MBTESTCD	MB.MBTEST	MB.MBCAT	MB.MBSCAT	MB.
MSSEQ	MS.MSGRPID	MS.MSREFID	MS.MSSPID	MS.MSTESTCD	MS.MSTEST	MS.MSCAT	MS.MSSCAT	MS.
ASEQ	FA.FAGRPID	FA.FAREFID	FA.FASPID	FA.FATESTCD	FA.FATEST	FA.FAOBJ	FA.FASCAT	FA.
IR	IDVARVAL	RELTYPE	RELID					
IR	IDVARVAL	QNAM	QLABEL	QVAL	QORIG	QEVAL		
UID	DM.RFSTDTCT	DM.RFENDTCT	DM.SITEID	DM.INVID	DM.INVNAM	DM.BRTHDTC	DM.AGE	DM.
DSSEQ	QS.QSGRPID	QS.QSSPID	QS.QSTESTCD	QS.QSTEST	QS.QSCAT	QS.QSSCAT	QS.QSORRES	QS.
VSIT	SV.VISITNUM	SV.VISITDY	SV.SVSTDTCT	SV.SVENDTCT	SV.SVUPDES			
ESEQ	PE.PEGRPID	PE.PESPID	PE.PETESTCD	PE.PETEST	PE.PEMODIFY	PE.PECAT	PE.PESCAT	PE.
ESEQ	AE.AEGRPID	AE.AEREFID	AE.AESPID	AE.AETERM	AE.AEMODIFY	AE.AEDECOD	AE.AECAT	AE.
VSSEQ	VS.VSGRPID	VS.VSSPID	VS.VSTESTCD	VS.VSTEST	VS.VSCAT	VS.VSSCAT	VS.VSPOS	VS.V

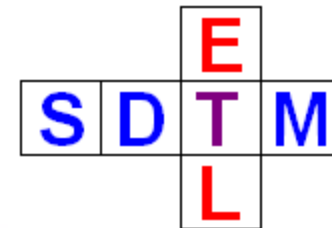
[www.xml4pharma.com/SDTM-ETL](http://www.xml4pharma.com/SDTM-ETL)



# CORE Implementation in Software

## SDTM-ETL, a "smart mapping tool"

- When generating SDTM/SEND datasets, one immediately wants to validate against the standard
- with the possibility to use a **subset** of rules only
- with the possibility to **select a set** of SDTM/SEND files, rather than all
- In an extremely user-friendly way ...



[www.xml4pharma.com/SDTM-ETL](http://www.xml4pharma.com/SDTM-ETL)

# SDTM-ETL is now implementing CORE

← → ↻ [www.xml4pharma.com/SDTM-ETL/](http://www.xml4pharma.com/SDTM-ETL/)

## NEW: Version 4.2, coming very soon!

The SDTM-ETL™ software is considered to be the lowest clinical data in CDISC ODM format (most EDC systems do not). SDTM-ETL is completely "SAS®-free", i.e. unlike other software solutions.

[SDTM-ETL allows you to reuse mappings from other studies](#)

SDTM-ETL comes with an extremely user-friendly graphical user interface. Details are provided using intelligent wizards (no XML editing).

Many CROs and service providers have already discovered the benefits of SDTM-ETL.

## Latest version: SDTM-ETL 4.1

SDTM-ETL v.4.1 has full support for as well define.xml 2.10.

The most important documentation and a number of manuals are available:

- [As of v.4.2: Validating SDTM/SEND datasets using CORE](#)
- [SDTM-ETL 4.1 new features overview](#)
- [SDTM-ETL 4.0 new features overview](#)

## SDTM-ETL 4.2: Validating SDTM/SEND datasets Using CORE (CDISC Open Rules Engine)

Author: Jozef Aerts, XML4Pharma

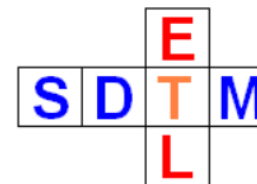
Last update: 2023-04-10

### Table of Contents

Introduction .....	1
Features of CORE .....	2
CORE Implementation in SDTM-ETL .....	2
Running CORE in SDTM-ETL .....	3
Conclusion .....	9

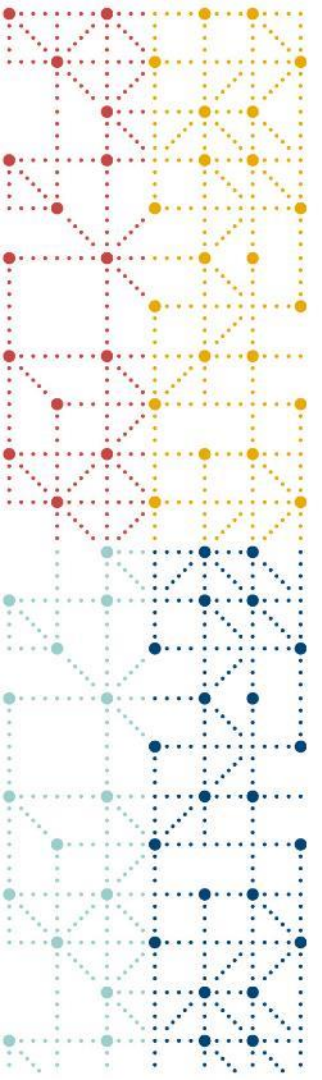
### Introduction

CDISC CORE (CDISC Open Rules Engine) is currently revolutionizing the submission validation world.





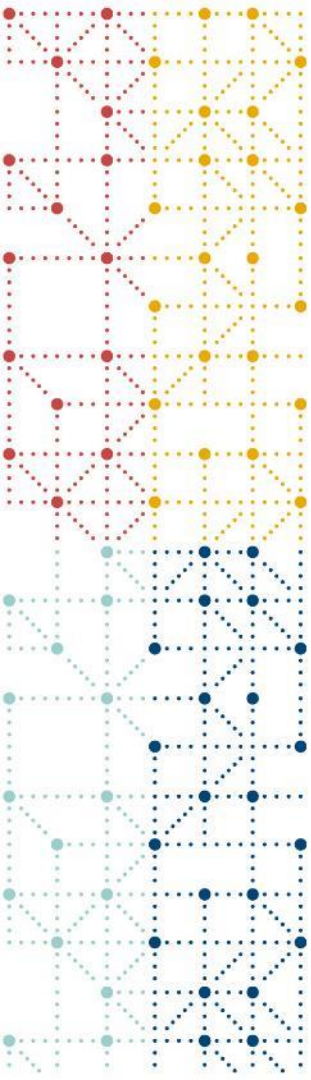




# Conclusions

# Conclusions

- CDISC CORE is easy to implement in software - just trigger a separate process
- As all rules and their implementations are open source and publicly available, we can finally get rid of "black box" software
- Not all rules have been implemented yet (but steadily growing)
- CORE is great for QC during the development of the mappings
- With CORE, every one "plays by the same rules"
- In future, CORE can be extended with company-internal rules, which makes it even more interesting



**Thank You!**

**Contact: [Jozef.Aerts@XML4Pharma.com](mailto:Jozef.Aerts@XML4Pharma.com)**

**cdisc**