



Health & Nutrition, Health Science

SI05: Developing and implementing CDISC CORE rules: from zero to hero

Health Inspired,
Quality Driven.

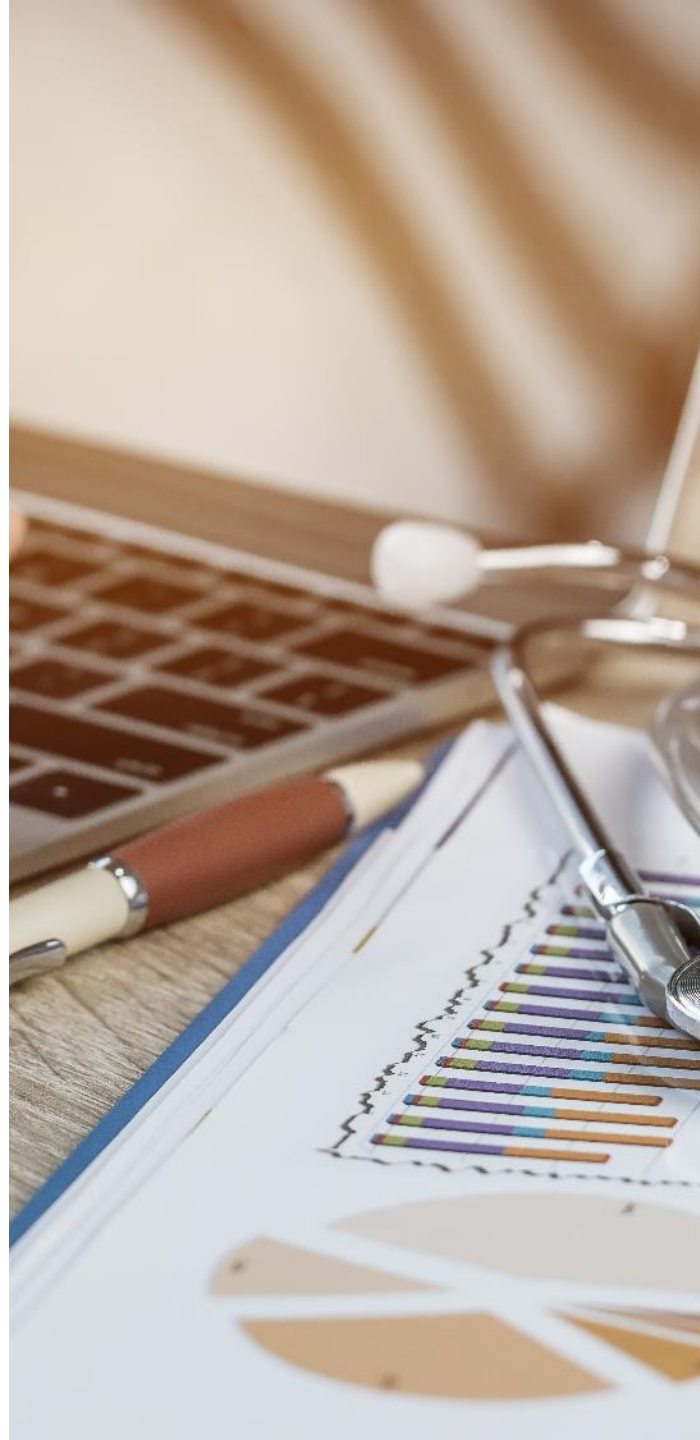


Disclaimer and disclosures

The views and opinions expressed in this presentation are those of the author(s) and do not necessarily reflect the opinions of PHUSE members, respective companies or organizations.

Agenda

- Introduction
- To CORE or not to CORE?
- Project Evolution
- HardCORE Practice
- Implementation
- Conclusion



01 Introduction



Introduction

- Global Mission
 - High quality data packages
 - Clear and fast analysis & reporting
 - Successful submission
- Reality
 - Increased study complexity
 - Several regulatory requirements
 - Flexibility and adaptability needed
 - Fast, faster, fastest



Introduction

- Global Mission
 - High quality data packages
 - Clear and fast analysis & reporting
 - Successful submission
- Reality
 - Increased study complexity
 - Several regulatory requirements
 - Flexibility and adaptability needed
 - Fast, faster, fastest
 - CDISC CORE





02 To CORE or not to CORE?

This should not be a question

To CORE or not to CORE?

- Company 
 - Daily validation of data packages
 - CDISC + regulatory + in-house developed rules
 - Rules that are up to date with latest standards
 - Rules that are unambiguous
 - Validation of only parts of the data (e.g. blinded data)
 - Achieve best quality in the most efficient way
 - Open Source

To CORE or not to CORE?

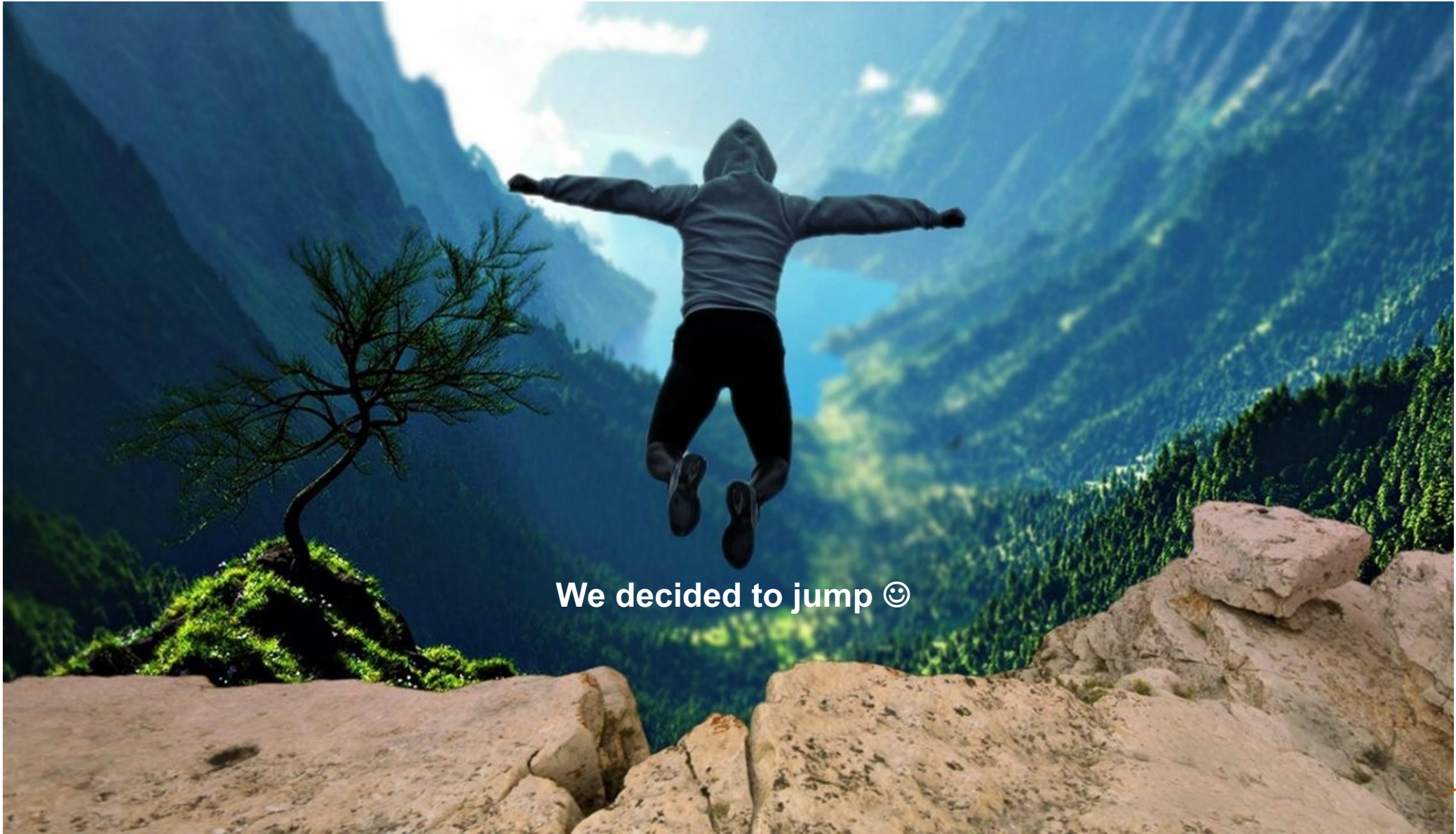
- Personal 
 - Expand knowledge of CDISC standards
 - Understand how rules are defined and developed @CDISC
 - Connect with people from the industry and exchange thoughts
 - Contribute to the development of the industry

To CORE or not to CORE?

- Professional experience as a Data Manager 
 - eCRF review
 - Data cleaning
 - SDTM set-up & validation
 - Development & validation of data cleaning rules (on SDTM)

- Cold Feet 
 - First time volunteering – What is expected from me?
 - Am I smart enough? – Imposter syndrome
 - What if they only need programmers? – I know some basic SQL but...

To CORE or not to CORE?



We decided to jump 😊

03 Project Evolution

Rules Development Team



Project Evolution

4 Project Pillars



Training

- Does everyone understand the process?
- Repeat trainings on a regular basis for newcomers
- Workshops with F2F assistance



Team Connection

- Regular meetings in an informal setting
- Everybody can ask questions + learn from each other
- Keep everyone up-to-date on new decisions



Information Retrieval

See next slides



Process Description and Conventions

See next slides

Project Evolution

Information Retrieval

■ Wiki pages

- ✓ Conformance Rule Development Team
 - Decision Log
 - › Executable Rule Authoring Processes
 - › Training
 - › **Reference Guide**
 - Rule Type Templates
 - CORE Executable Status
 - › Conformance Rule Development Team Meeting Notes
 - › **SDTMIG Conformance Rules Sign-up List**

■ **GitHub**: allows better tracking of scripting updates

Direct link Jira issue - Github issue

■ **Jira Tracker**: allows better follow-up of rule statuses

○	CORERULES-823	CG0066	Marisa Wyckmans i	Amy Palmer i	QC IN PROGRESS
○	CORERULES-4989	SD0024	Marisa Wyckmans i	Kelli Smith i	PUBLISHED
○	CORERULES-4913	SD0012	Marisa Wyckmans i	Kelli Smith i	PUBLISHED
○	CORERULES-4920	SD0013	Marisa Wyckmans i	Kelli Smith i	PUBLISHED
○	CORERULES-6918	SD2272	Marisa Wyckmans i	Kelli Smith i	AUTHOR IN PROGRESS

Conformance Rules Editor

Reference Guide

Boolean Logic

Check Operator

Check Parameter

Match Datasets (Merge)

Operations (Aggregate)

Rule Type

Scope

Sensitivity

Metadata Variables

Test Data

Test Results

Ticket Submission

FAQ

Project Evolution

Process Description and Conventions

- Good documentation of different steps:
 - Early day volunteers versus people joining later
 - Small group versus larger group
 - Fixed workflow



1 – Write Rule

- Source
- Template = scheme
- Operators
- Rule logic

2 – Create Test Data

- Format
- Examples
- + and –
- Location

3 – QC Rule Description

- = Source
- Rule logic
- Syntax
- Scheme

4 – Validate Rule Output

- Format
- Test data
- Location

5 – Publish Rule



Feedback to author



04 HardCORE practice

How to develop and validate rules

HardCORE Practice #1

- Source = SDTM and SDTMIG Conformance Rules v2.0

Rule ID	SDTMIG Version	Rule Version	Class	Domain	Variable	Condition	Rule
CG0344	3.4	1	FND	ALL	--TESTCD	--TESTCD ^= 'MULTIPLE'	

- Rule Editor: simple YAML syntax
- Written to output in case data is not conform the rule
→ detect errors in data
- Clear message to indicate what is incorrect in your data

```
72 Scope:
73   Classes:
74     Include:
75       - FINDINGS
76   Domains:
77     Include:
78       - ALL
79
```

```
EDIT TEST
1 # Variable: --TESTCD
2 # Condition:
3 # Rule: --TESTCD ^= 'MULTIPLE'
4 Check:
5   all:
6     - name: --TESTCD
7       operator: equal_to
8       value: MULTIPLE
9 Core:
10   Id: CORE-000162
11   Version: '1'
12   Status: Draft
13   Description: Raise an error when --TESTCD = 'MULTIPLE'
14   Outcome:
15     Message: --TESTCD is equal to 'MULTIPLE'
16     Output Variables:
17       - --TESTCD
18   Rule Type: Record Data
19   Sensitivity: Record
20   Executability: Fully Executable
21   Authorities:
22     Organization: CDISC
23     Standards:
24       - Name: SDTMIG
25         Version: '3.4'
26         References:
27           - Origin: SDTM and SDTMIG Conformance Rules
28             Rule Identifier:
29               Id: CG0344
30               Version: '1'
31               Version: '2.0'
32         Citations:
33           - Cited Guidance: If multiple result values (--ORRES) are reported
34             for
35               a test in a Findings class dataset, multiple records should be
36               submitted
37               for that --TESTCD.
38             Document: IG v3.4
39             Section: 4.2.8.2
```

HardCORE Practice #2

- Source = SDTM and SDTMIG Conformance Rules v2.0

Rule ID	SDTMIG Version	Rule Version	Class	Domain	Variable	Condition	Rule
CG0066	3.2	1	EVT	DS	DSDECOD	DSCAT = 'PROTOCOL MILESTONE'	DSTERM = DSDECOD

- Condition:
 - Only subset of data is in scope
 - Added to rule logic in rule editor

EDIT

TEST

```
22 Core:
23   Id: CDISC.SDTMIG.CG0066
24   Version: '1'
25   Status: Draft
26   Description: 'When DSCAT equals PROTOCOL MILESTONE, then DSTERM should be equal to DSDECOD'
27   Outcome:
28     Message: 'DSCAT equals PROTOCOL MILESTONE but DSTERM is not equal to DSDECOD'
29     Output Variables:
30       - DSTERM
31       - DSDECOD
32   Rule Type: Record Data
33   Sensitivity: Record
34   Scope:
35     Classes:
36       Include:
37         - EVENTS
38     Domains:
39       Include:
40         - DS
41   Executability: Fully Executable
42   Check:
43     all:
44       - name: DSCAT
45         operator: equal_to
46         value: PROTOCOL MILESTONE
47         value is literal: true
48       - name: DSTERM
49         operator: not_equal_to
50         value: DSDECOD
```

HardCORE Practice #2

Validation in Rule Editor

Does everything work?

EDIT

TEST

✓

Validate YAML Syntax

✓

Validate YAML against Schema

✓

Convert YAML to JSON Rule

⌚

Load Test Define.xml

⌚

Load Test Datasets

⌚

Results

Loading test data

✓ Load Test Datasets

TEST DATASETS FILE...

Filename: unit-test-coreid-CG0066-negative.xlsx

USUBJID	DSTERM	DSDECOD	DSCAT
	Reported Term for the Disposition Event	Reported Term for the Disposition Event	Reported Term for the Disposition
Unique Subject Identifier			
Char	Char	Char	Char
21	200	200	200
015246-076-0003-00003	FUTURE USE OF SAMPLES INFORMED CONSENT OBTAINED	Future USE OF SAMPLES INFORMED CONSENT OBTAINED	PROTOCOL MILESTONE
015246-203-0003-00001	STUDY INFORMED CONSENT OBTAINED	STUDY INFORMED CONSENT OBTAINED	PROTOCOL MILESTONE
015246-300-0004-00002	COMPLETED	ADVERSE EVENT	PROTOCOL MILESTONE

Check output – as expected?

✓ Results **Negatives** 3

Results

```
{ 1 item
  "DS": [ 1 item
    0: { 5 items
      "executionStatus": "success"
      "domain": "DS"
      "variables": [ 2 items
        0: "DSDECOD"
        1: "DSTERM"
      ]
      "message":
        "DSCAT equals PROTOCOL MILESTONE but DSTERM is not equal to DSDECOD"
      "errors": [ 3 items
        0: { 3 items
          "value": { 2 items
            "DSDECOD":
              "Future USE OF SAMPLES INFORMED CONSENT OBTAINED"
            "DSTERM":
              "FUTURE USE OF SAMPLES INFORMED CONSENT OBTAINED"
          }
          "row": 1
          "USUBJID": "015246-076-0003-00003"
        }
      ]
    }
  ]
}
```

HardCORE Practice #3

- Source = SDTM and SDTMIG Conformance Rules v2.0

Rule ID	SDTMIG Version	Rule Version	Class	Domain	Variable	Condition	Rule
CG0148	3.4	2	SPC	DM	RFXSTDTC	EX records present for subject	RFXSTDTC = earliest EX.EXSTDTC

- Operations to define 'new' values for use in rule logic:
 - Distinct
 - Min, Max
 - And many more in reference guide

EDIT TEST

```
1 # Variable: RFXSTDTC
2 # Condition: EX records present for subject
3 # Rule: RFXSTDTC = earliest EX.EXSTDTC
4 Check:
5   all:
6     - name: USUBJID
7       operator: is_contained_by
8       value: $ex_usubjid
9     - name: RFXSTDTC
10      operator: not_equal_to
11      value: $min_ex_exstdtc
12 Core:
13   Id: CORE-000239
14   Version: '1'
15   Status: Published
16   Description: Raise an error when EX records are present for subject but RFXSTDTC
17               does
18               not equal the earliest value of EX.EXSTDTC
19 Operations:
20   - operator: distinct
21     domain: EX
22     name: USUBJID
23     id: $ex_usubjid
24   - operator: min_date
25     domain: EX
26     name: EXSTDTC
27     id: $min_ex_exstdtc
28     group:
29       - USUBJID
30 Outcome:
31   Message: RFXSTDTC does not equal the earliest value of EX.EXSTDTC
32   Output Variables:
33     - RFXSTDTC
34     - $min_ex_exstdtc
35 Rule Type: Record Data
36 Sensitivity: Record
37 Executability: Fully Executable
```

HardCORE Practice #4

- QC of rule logic in rule editor
 - Yellow squiggly lines = automatic detection of errors

- Validation error

EDIT	TEST
✓	Validate YAML Syntax
✗	Validate YAML against Schema
✓	Convert YAML to JSON Rule

```
Check:
all:
  - name: TSPARMCD
    operator: not_contains
    value: OBJSEC
    value_is_literal: true
Core:
  Id: FDA.SDTMIG.CG0700
  Status: Draft
  Version: '1'
Description: "'Trial Secondary Objective' (OBJSEC) is an 'FDA Desired' parameter listed in the Technical Conformance Guide (TCG). OBJSEC record must be present in Trial Summary (TS) domain."
Executability: Fully Executable
Outcome:
  Message: Missing OBJSEC Trial Summary Parameter
  Output Variables:
    - STUDYID
Rule Type: Record Data
Scope:
  Classes:
    Include:
      - TRIAL DESIGN
  Domains:
    Include:
      - TS
Sensitivity: Dataset
```

HardCORE Practice #4

- QC of rule logic in rule editor
 - Yellow squiggly lines = automatic detection of errors
- Update rule logic
- Create test data

```
Check:
  all:
    - name: TSPARMCD
      operator: not_contains
      value: OBJ
      value_is_literal: true
      Value must be "additional_columns_empty" |
      "additional_columns_not_empty" | "conformant_value_data_type" |
      "conformant_value_length" | "contains" | "contains_all" |
      "contains_case_insensitive" | "date_equal_to" |
      "date_greater_than" | "date_greater_than_or_equal_to" |
      "date_less_than" | "date_less_than_or_equal_to" |
      "date_not_equal_to" | "does_not_contain" |
      "does_not_contain_case_insensitive" |
      "does_not_equal_string_part" |
      "does_not_have_next_corresponding_record" |
      "does_not_reference_correct_codelist" |
      "does_not_use_valid_codelist_terms" | "empty" |
      "empty_within_except_last_row" | "ends_with" | "equal_to" |

Core:
  Id: FDA.SDTMIG
  Status: Draft
  Version: '1'
  Description: "'T
parameter listed
must be present
Executability: F
Outcome:
  Message: Missi
  Output Variabl
    - STUDYID
  Rule Type: Record Data
  Scope:
    Classes:
      Include:
        - TRIAL DESIGN
```

```
Check:
  all:
    - name: TSPARMCD
      operator: does_not_contain
      value: OBJSEC
      value_is_literal: true
```

HardCORE Practice #4

- Source = FDA Business Rules v1.5

Version 1.5 finalized June 2019

FDA Business Rule ID	FDA Business Rule
Clinical and Nonclinical	
FDAB011	All trial design data should be submitted as specified in the FDA Study Data Technical Conformance Guide (TCG).

- TCG – Appendix B

FDA Desired - Clinical	TSPARMCD	TSPARM	FDA Notes
Y	OBJSEC	Trial Secondary Objective	Use as many rows as needed.

- QC of rule logic in rule editor
- Create test data
- Update, validate & publish rule

```
- Name: SDTMIG
References:
  - Citations:
    - Cited Guidance: FDAB011
      Document: FDA
      Origin: FDA Business Rules
      Rule Identifier:
        Id: CG0700
        Version: '1'
        Version: '1.5'
        Version: '3.3'

Check:
  all:
    - name: $distinct_tsparmcd
      operator: does_not_contain
      value: OBJSEC
      value_is_literal: true

Core:
  Id: FDA.SDTMIG.CG0700
  Status: Draft
  Version: '1'
  Description: "'Trial Secondary Objective' (OBJSEC) is an 'FDA Desired' parameter listed in the Technical Conformance Guide (TCG). OBJSEC record must be present in Trial Summary (TS) domain."
  Executability: Fully Executable
  Operations:
    - id: $distinct_tsparmcd
      name: TSPARMCD
      operator: distinct

Outcome:
  Message: Missing OBJSEC Trial Summary Parameter
  Output Variables:
    - STUDYID

Rule Type: Record Data
Scope:
  Classes:
    Include:
      - TRIAL DESIGN
  Domains:
    Include:
      - TS
  Sensitivity: Dataset
```

HardCORE Practice

Key Learnings



Achievements

- Rule writing & validation is easy and very accessible – no cold feet needed
- One requirement: knowledge of the data standard – no programming
- Amazing work behind the scenes to translate simple rule code into complex scripts
- Hard work from project leads and volunteers – hint: bring a friend



Challenges

- Impossible to foresee everything upfront – develop while doing – rework needed
- Will everyone do it correctly? – control
- Non-executable or ambiguous rules
- Keep info up to date and remove obsolete information – avoid confusion

05 Implementation



Implementation

- Ongoing:
 - CLI of the CORE engine allows for an easy integration in our daily conversion process
 - Evaluate this integration by running the current available rules in the background
- Future:
 - Use of CDISC CORE in production → outputs refreshed daily
 - Validation of data packages via CORE
- Exploratory:
 - Deploy a local installation of the CDISC rule editor in the SGS cloud
 - Goal: to create and run custom rules



06 Conclusion

We took the decision to jump ...

It was the right decision, we survived ...

So no excuses, just jump ...





Questions?

Visit us at booth 32

Contact us

clinicalresearch@sgs.com

EUROPE: +32 15 27 32 45

AMERICAS: + 1 877 677 2667

www.sgs.com/cro

health-nutrition@sgs.com
www.sgs.com/healthnutrition

