Health & Nutrition, Health Science

Developing and implementing CDISC CORE rules: from zero to hero

Health Inspired,
Quality Driven.

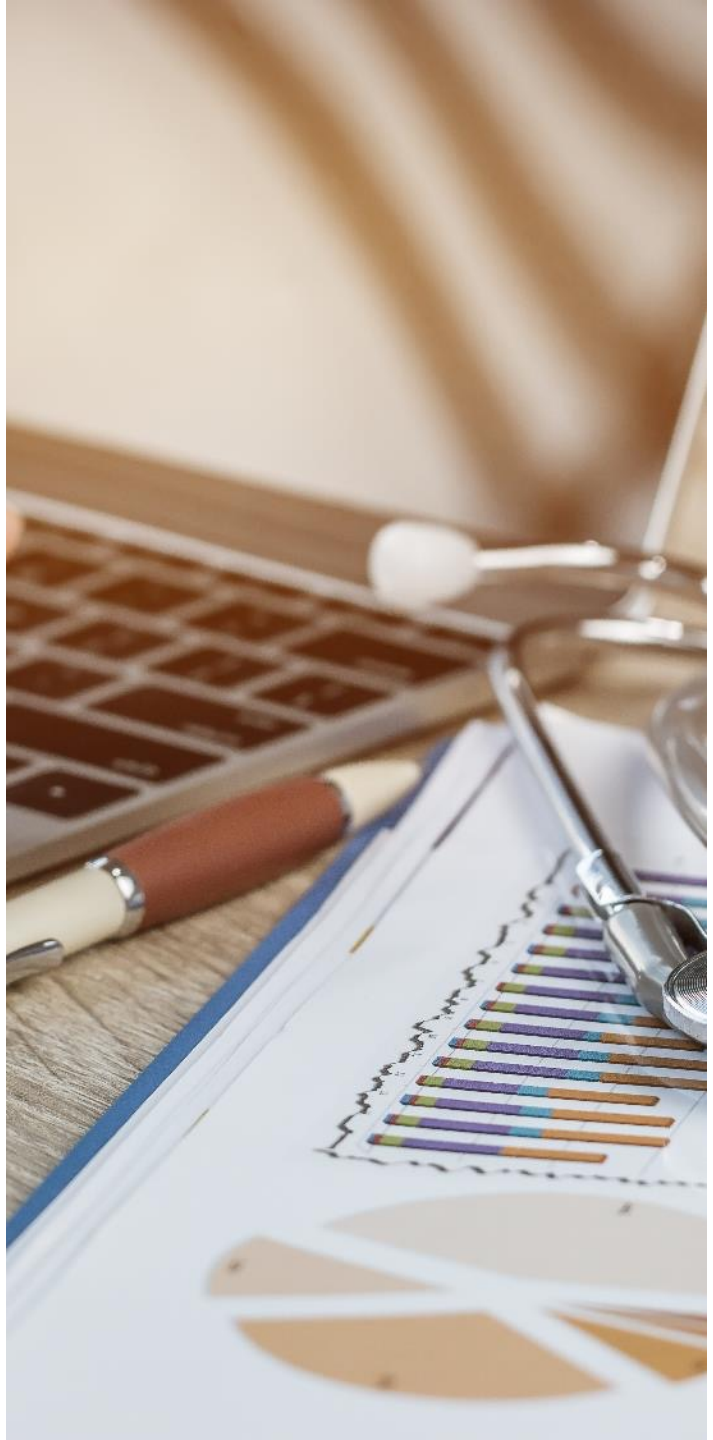
ACDM 2024 | Copenhagen | March 4th, 2024
Marisa Wyckmans – Data Management System and Process Manager

SGS

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 ACDM members, respective companies or organizations.

Agenda



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

01 Introduction & CORE Concept

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

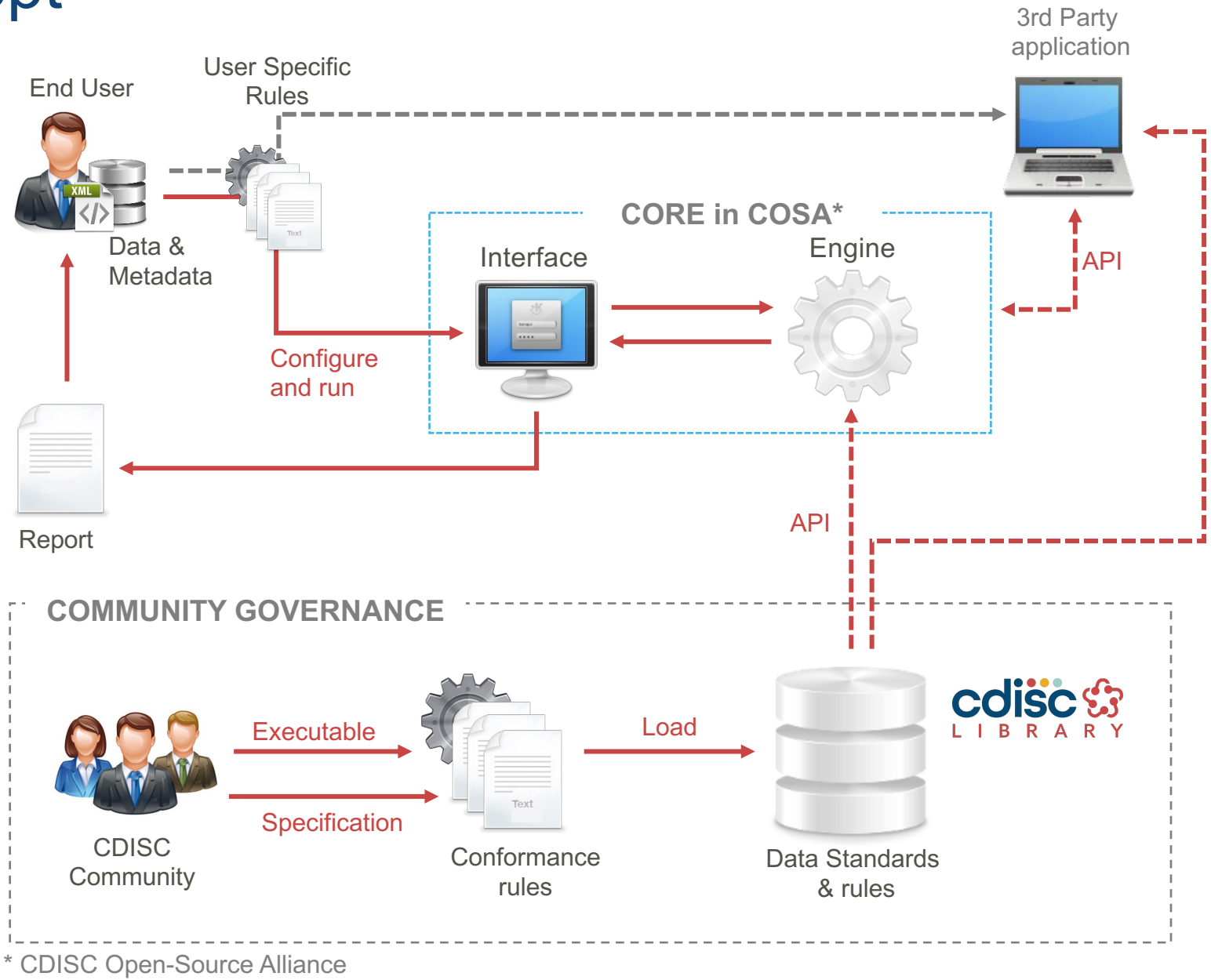
Introduction

- 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

- Conformance Rules, CDISC Open Rules Engine, Peter Van Reusel & Amy Palmer, CDISC US Interchange 2023
- CDISC Conformance rules and the CDISC Open Rules Engine: Continuing the Road to Adoption, Nick De Donder, PHUSE EU Connect 2023

CORE Concept



Volume and Breadth of Conformance Rules

Over 1900 Unique Conformance Rules

Covers multiple different versions of the Implementation Guides

Includes Regulatory Rules

SDTMIG v3.2, v3.3, v3.4

ADaMIG v1.0, v1.1, v1.2, v1.3, PopPK

SENDIG v3.0, v3.1, v3.1.1, AR v1.0, DART, GeneTox

- Conformance Rules, CDISC Open Rules Engine, Peter Van Reusel & Amy Palmer, CDISC US Interchange 2023
- CDISC Conformance rules and the CDISC Open Rules Engine: Continuing the Road to Adoption, Nick De Donder, PHUSE EU Connect 2023

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?

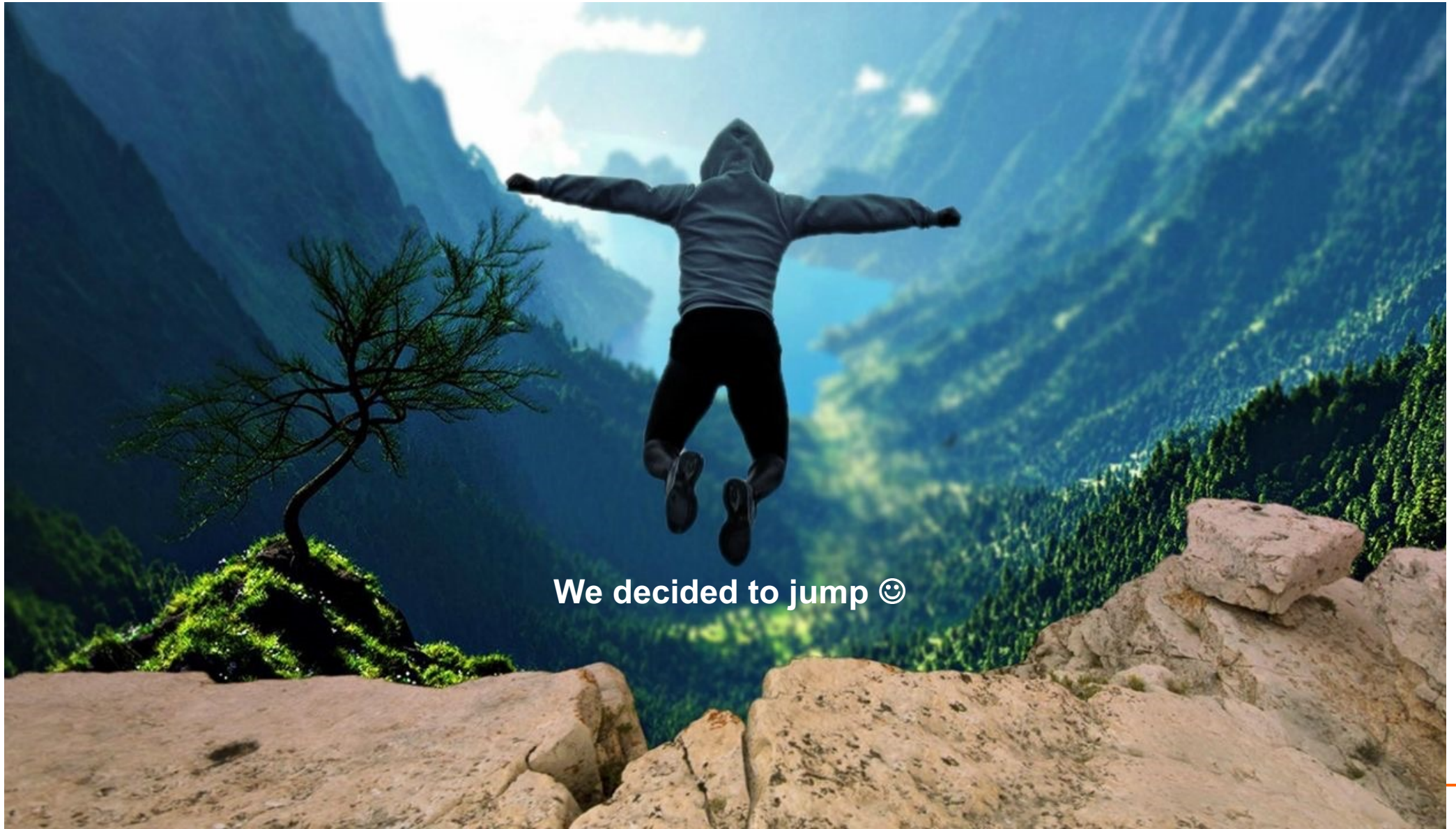
- Personal 
  - Expand knowledge of CDISC standards
  - Understand how rules are defined and developed @CDISC
  - Connect with people from the industry and exchange thoughts
  - Contribute to the development of the industry

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?



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

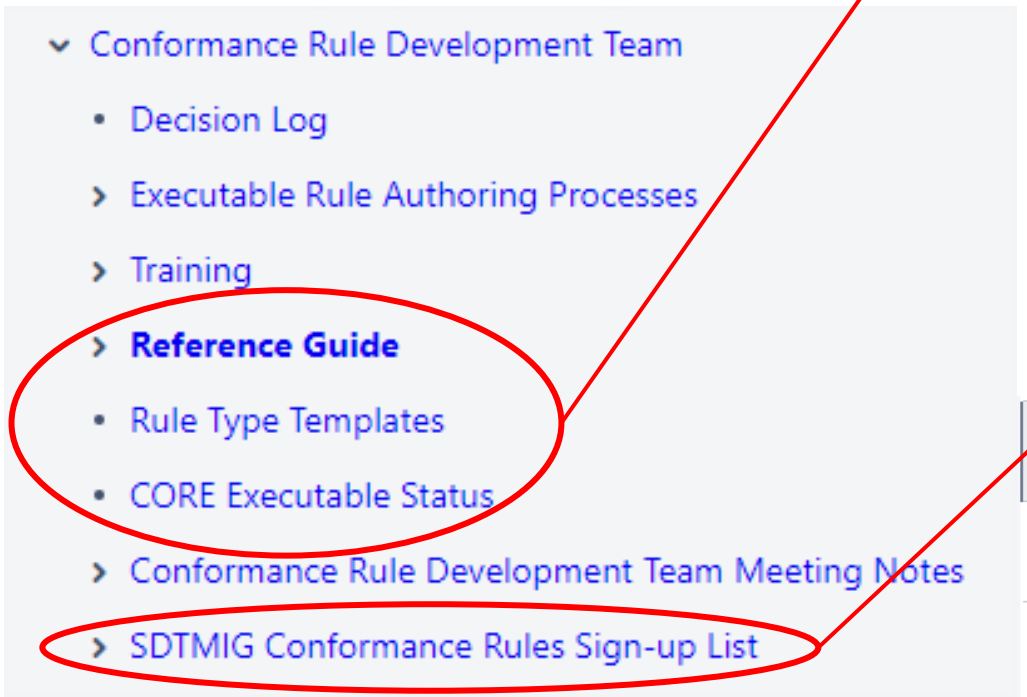
Process Description and Conventions

See next slides

Project Evolution

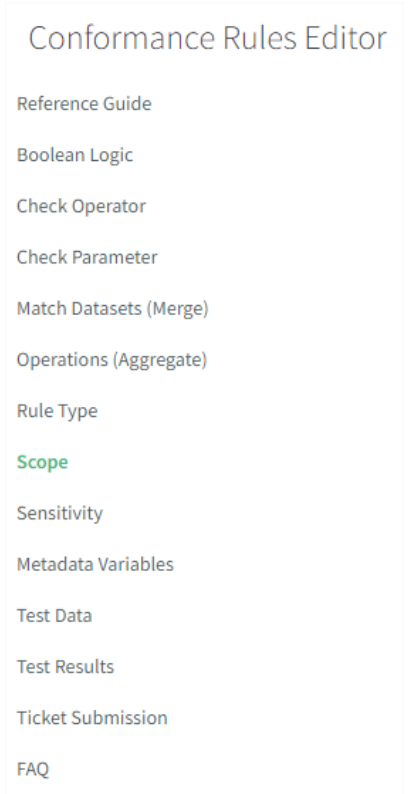
Information Retrieval

- Wiki pages



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

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

EDIT TEST

Automatic detection of errors

```yaml
# Variable: --TESTCD
# Condition:
# Rule: --TESTCD ^= 'MULTIPLE'
Check:
  all:
    - name: --TESTCD
      operator: is_equal_to
      value: MULTIPLE
Core:
  Id: CORE-000162
  Version: '1'
  Status: Draft
Description: Raise an error when --TESTCD = 'MULTIPLE'
Outcome:
  Message: --TESTCD is equal to 'MULTIPLE'
  Output Variables:
    - --TESTCD
Rule Type: Record Data
Sensitivity: Record
Executability: Fully Executable
Authorities:
  Organization: CDISC
  Standards:
    - Name: SDTMIG
      Version: '3.4'
      References:
        - Origin: SDTM and SDTMIG Conformance Rules
          Rule Identifier:
            Id: CG0344
            Version: '1'
          Version: '2.0'
          Citations:
            - Cited Guidance: If multiple result values (--ORRES) are reported
              for
                a test in a Findings class dataset, multiple records should be
                submitted
                for that --TESTCD.
              Document: IG v3.4
              Section: 4.2.8.2
Scope:
  Classes:
    Include:
      - FINDINGS
  Domains:
    Include:
      - ALL
```

```yaml
- name: --TESTCD
  operator: equal_to
  value: MULTIPLE
```

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
Unique Subject Identifier	*Reported Term for the Disposition Event*	*Reported Term for the Disposition Event*	*Reported Term for the Disposition*
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 OBTAINE	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

```yaml
# Variable: RFXSTDTC
# Condition: EX records present for subject
# Rule: RFXSTDTC = earliest EX.EXSTDTC
Check:
  all:
    - name: USUBJID
      operator: is_contained_by
      value: $ex_usubjid
    - name: RFXSTDTC
      operator: not_equal_to
      value: $min_ex_exstdtc
Core:
  Id: CORE-000239
  Version: '1'
  Status: Published
Description: Raise an error when EX records are present for subject but RFXSTDTC does
  not equal the earliest value of EX.EXSTDTC
Operations:
  - operator: distinct
    domain: EX
    name: USUBJID
    id: $ex_usubjid
  - operator: min_date
    domain: EX
    name: EXSTDTC
    id: $min_ex_exstdtc
    group:
      - USUBJID
Outcome:
  Message: RFXSTDTC does not equal the earliest value of EX.EXSTDTC
  Output Variables:
    - RFXSTDTC
    - $min_ex_exstdtc
Rule Type: Record Data
Sensitivity: Record
Executability: Fully Executable
```

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

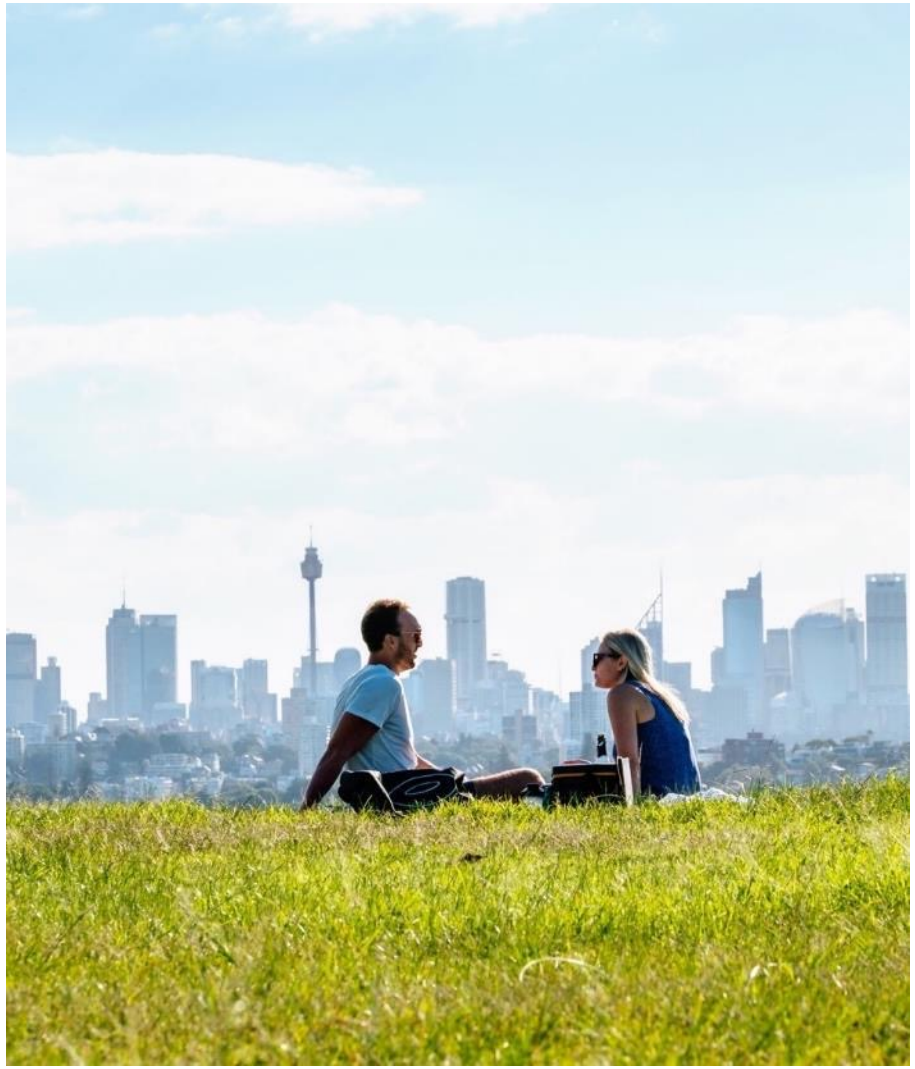
- Exploratory:
  - Deploy a local installation of the CDISC rule editor in the SGS cloud
  - Goal: to create and run custom rules

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 36

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