



When you need
to be sure

SGS Health Science

CDISC Open Rules: a deep dive into custom rule creation

Roman Radelicki, Head Data Technology, PHUSE US Connect 2025



Speaker

Roman Radelicki

Head Data Technology

SGS Health Science

clinicalresearch@sgs.com

Roman Radelicki started his career as Programmer in 2006 and joined SGS in 2009. During his career at SGS he held several positions and became Head Data Technology in 2021. He is mainly responsible for managing the data engineering and application development teams, providing support to the different departments



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 PHUSE and/or SGS.*
- *The author(s) have no real or apparent conflicts of interest to report.*

Agenda



- 01 About CDISC Open Rules
- 02 Local custom rules
- 03 Who can create rules and why should we
- 04 Creating custom CDISC Open Rules
- 05 Use cases
- 06 Suggestions – AI integration
- 07 Key takeaways
- 08 Q&A

Agenda



01 About CDISC Open Rules

02 Local custom rules

03 Who can create rules and why should we

04 Creating custom CDISC Open Rules

05 Use cases

06 Suggestions – AI integration

07 Key takeaways

08 Q&A

About CDISC Open Rules

Rules

- CDISC governed – single source of truth
- Community driven
- Executable rules
- Submission ready
- Open-Source



Agenda



01 About CDISC Open Rules

02 Local custom rules

03 Who can create rules and why should we

04 Creating custom CDISC Open Rules

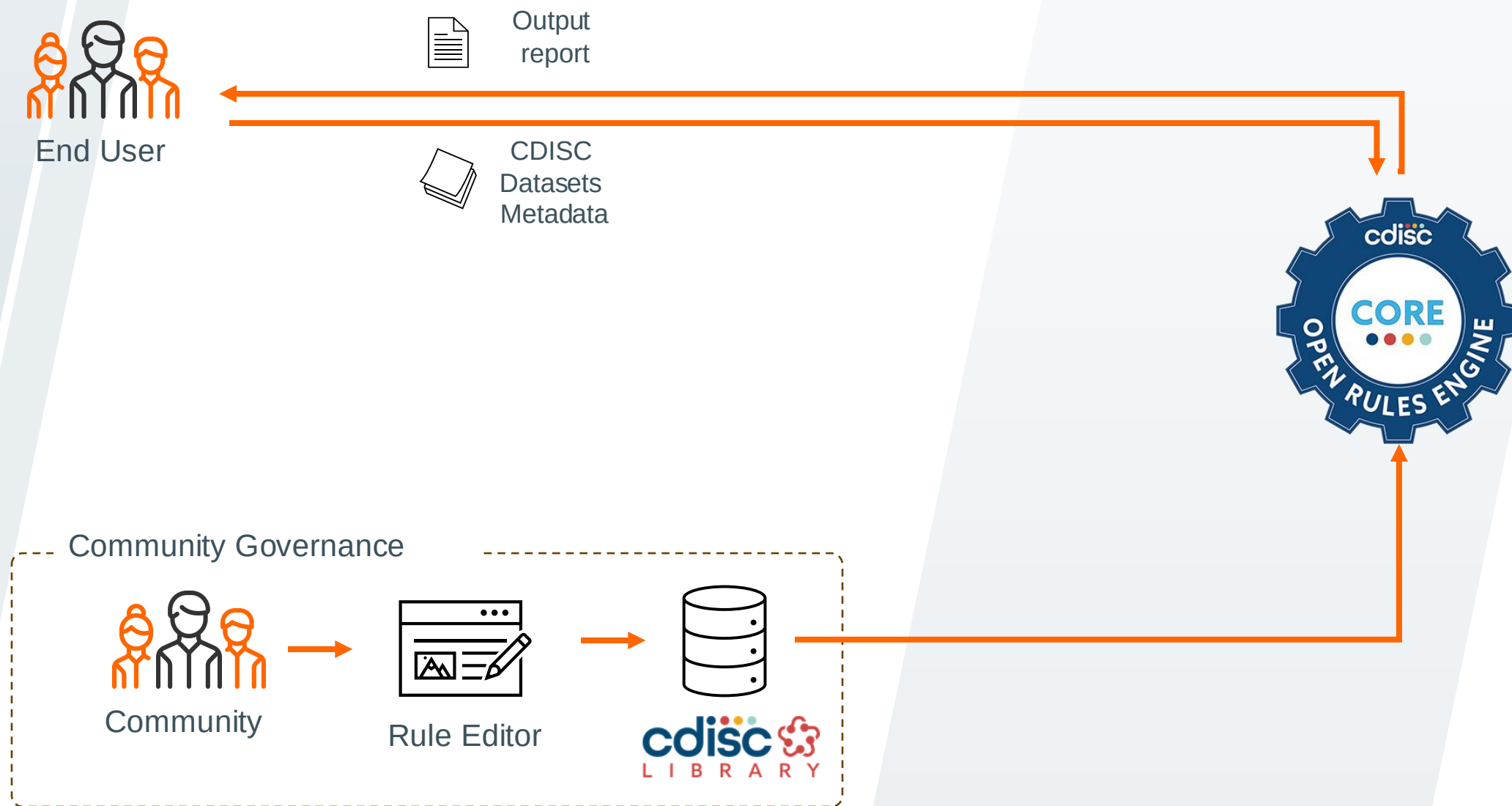
05 Use cases

06 Suggestions – AI integration

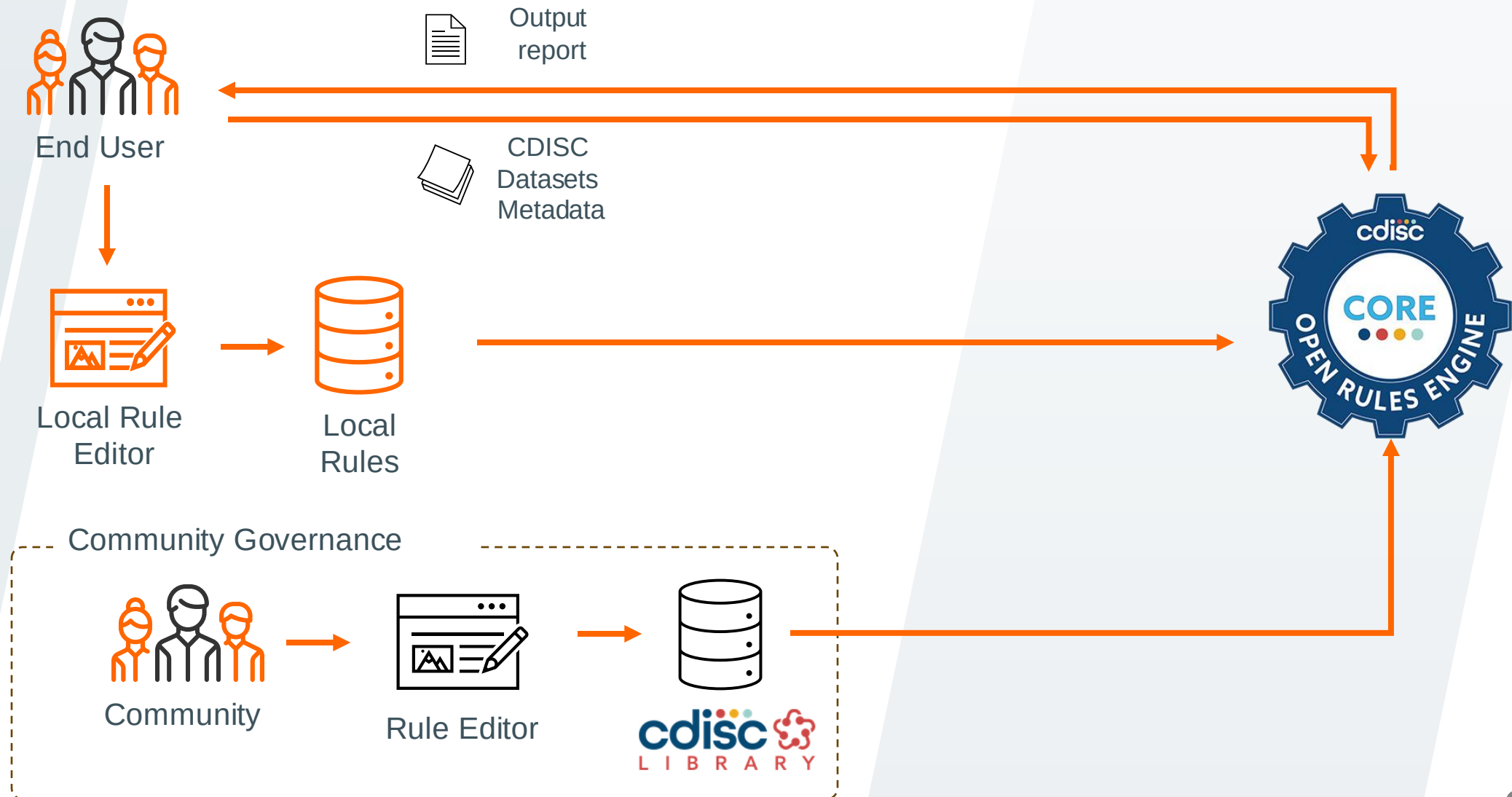
07 Key takeaways

08 Q&A

Local custom rules



Local custom rules



Agenda



01 About CDISC Open Rules

02 Local custom rules

03 Who can create rules and why should we

04 Creating custom CDISC Open Rules

05 Use cases

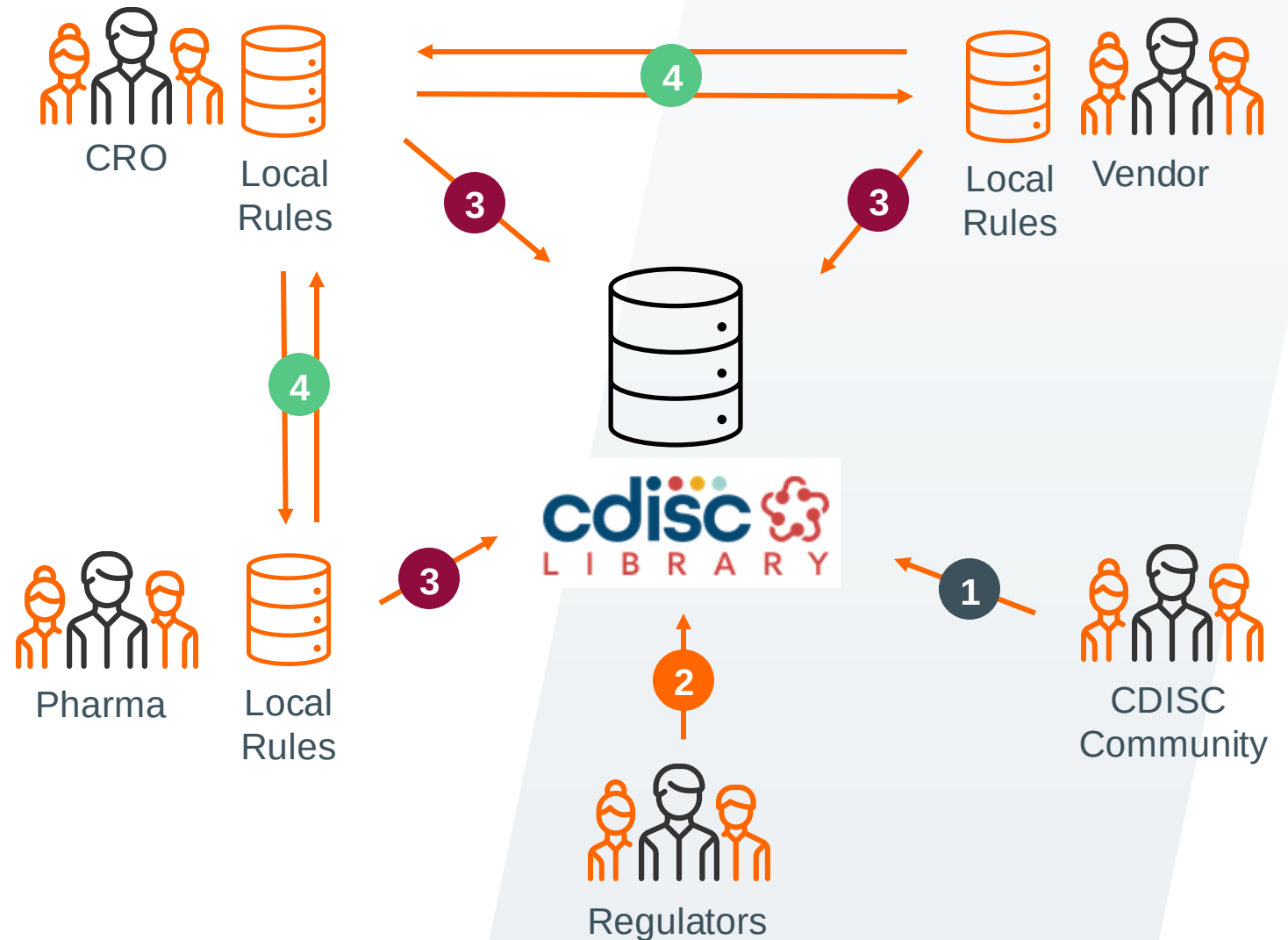
06 Suggestions – AI integration

07 Key takeaways

08 Q&A

Who & Why?

- 1 CDISC rules governance
- 2 Collaboration CDISC / FDA
- 3 Contribution to CDISC rules
- 4 Transfer of custom rules



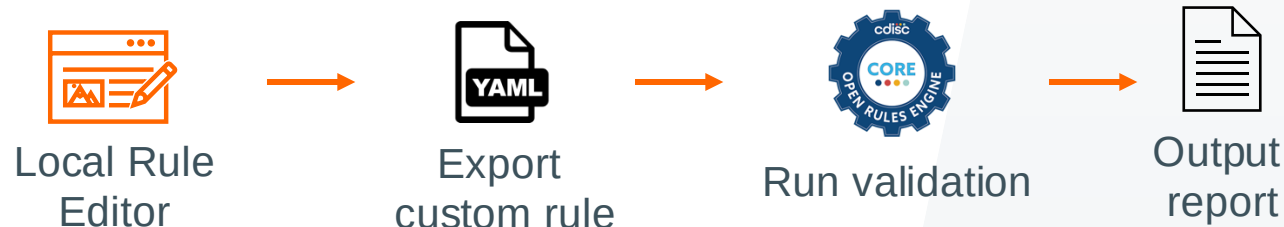
Agenda



- 01 About CDISC Open Rules
- 02 Local custom rules
- 03 Who can create rules and why should we
- 04 Creating custom CDISC Open Rules**
- 05 Use cases
- 06 Suggestions – AI integration
- 07 Key takeaways
- 08 Q&A

Creating custom CDISC Open Rules

Process flow:



Rule Editor

- Web application
- Written in TypeScript
- YAML
- Real-time syntax checking
- GitHub workflow provided

Rule Engine

- Download from GitHub
- Command line interface
 - integrate in existing process flows e.g. Nightly automatic conversion workflow
- CLI – UI interface for less tech-savvy people – SGS proof of concept



Agenda



- 01 About CDISC Open Rules
- 02 Local custom rules
- 03 Who can create rules and why should we
- 04 Creating custom CDISC Open Rules
- 05 Use cases**
- 06 Suggestions – AI integration
- 07 Key takeaways
- 08 Q&A

Use case: data listings

- List all male subjects older than 40 and female subjects older than 41

```
Check:
  any:
    - all:
      - name: AGE
        operator: greater_than
        value: 40
      - name: SEX
        operator: equal_to
        value: M
    - all:
      - name: AGE
        operator: greater_than
        value: 41
      - name: SEX
        operator: equal_to
        value: F
```

```
48 ✓ Check:
49 ✓   any:
50 ✓     - all:
51 ✓       - name: AGE
52           operator: greater_than
53           value: 40
54 ✓       - name: SEX
55           operator: equal_to
56           value: M
57 ✓     - all:
58 ✓       - name: AGE
59           operator: greater_than
60           value: 41
61 ✓       - name: SEX
62           operator: equal_to
63           value: F
64 ✓ Core:
65   Id: "ROMAN-0002"
66   Status: Draft
67   Version: '1'
68   Description: in Accordance to the inclusion/exclusion criteria, Raise an error if
69                 the male subject's age is greater than 40 and the female subject's age is greater
70                 than 41.
69   Executability: Fully Executable
70 ✓ Outcome:
71   Message: AGE greater than 40 and SEX is equal to "M" or AGE greater than 41 and
72           SEX is equal to "F".
72   Rule Type: Record Data
73 ✓ Scope:
74 ✓   Classes:
75 ✓     Include:
76         - SPECIAL PURPOSE
77 ✓   Domains:
78 ✓     Include:
79         - DM
80   Sensitivity: Record
```

Use case: data listings

CORE-ID	Message	Dataset	USUBJID	Record	Variable(s)	Value(s)
ROMAN-0002	AGE greater than 40 and SEX is equal to "M" or AGE greater than 41 and SEX is equal to "F".	DM	SGS-DRG-001-01-S001	1	AGE, SEX	42.0, F
ROMAN-0002	AGE greater than 40 and SEX is equal to "M" or AGE greater than 41 and SEX is equal to "F".	DM	SGS-DRG-001-01-S014	7	AGE, SEX	42.0, M
ROMAN-0002	AGE greater than 40 and SEX is equal to "M" or AGE greater than 41 and SEX is equal to "F".	DM	SGS-DRG-001-01-S024	9	AGE, SEX	48.0, M
ROMAN-0002	AGE greater than 40 and SEX is equal to "M" or AGE greater than 41 and SEX is equal to "F".	DM	SGS-DRG-001-01-S033	13	AGE, SEX	45.0, F
ROMAN-0002	AGE greater than 40 and SEX is equal to "M" or AGE greater than 41 and SEX is equal to "F".	DM	SGS-DRG-001-01-S034	14	AGE, SEX	55.0, F
ROMAN-0002	AGE greater than 40 and SEX is equal to "M" or AGE greater than 41 and SEX is equal to "F".	DM	SGS-DRG-001-01-S039	16	AGE, SEX	41.0, M
ROMAN-0002	AGE greater than 40 and SEX is equal to "M" or AGE greater than 41 and SEX is equal to "F".	DM	SGS-DRG-001-01-S047	19	AGE, SEX	48.0, F
ROMAN-0002	AGE greater than 40 and SEX is equal to "M" or AGE greater than 41 and SEX is equal to "F".	DM	SGS-DRG-001-01-S048	20	AGE, SEX	44.0, M
ROMAN-0002	AGE greater than 40 and SEX is equal to "M" or AGE greater than 41 and SEX is equal to "F".	DM	SGS-DRG-001-01-S049	21	AGE, SEX	43.0, M
ROMAN-0002	AGE greater than 40 and SEX is equal to "M" or AGE greater than 41 and SEX is equal to "F".	DM	SGS-DRG-001-01-S052	23	AGE, SEX	51.0, M

Use case: data listings

- Current output file/structure not ideal
- Indications

USUBJID	AGE	SEX	IND_COUNTRY_BEL
SGS-DRG-00A-S001	42	F	*
SGS-DRG-00A-S002	42	M	
SGS-DRG-00A-S003	48	M	
SGS-DRG-00A-S004	55	F	*
SGS-DRG-00A-S005	45	F	
SGS-DRG-00A-S006	41	M	
SGS-DRG-00A-S007	51	M	*

CORE-ID	Message	Dataset	USUBJID	Record	Variable(s)	Value(s)
ROMAN-0002	AGE greater than 40 and SEX is equal to "M" or AGE greater than 41 and SEX is equal to "F".	DM	SGS-DRG-001-01-S001	1	AGE, SEX	42.0, F
ROMAN-0002	AGE greater than 40 and SEX is equal to "M" or AGE greater than 41 and SEX is equal to "F".	DM	SGS-DRG-001-01-S014	7	AGE, SEX	42.0, M
ROMAN-0002	AGE greater than 40 and SEX is equal to "M" or AGE greater than 41 and SEX is equal to "F".	DM	SGS-DRG-001-01-S024	9	AGE, SEX	48.0, M
ROMAN-0002	AGE greater than 40 and SEX is equal to "M" or AGE greater than 41 and SEX is equal to "F".	DM	SGS-DRG-001-01-S033	13	AGE, SEX	45.0, F
ROMAN-0002	AGE greater than 40 and SEX is equal to "M" or AGE greater than 41 and SEX is equal to "F".	DM	SGS-DRG-001-01-S034	14	AGE, SEX	55.0, F
ROMAN-0002	AGE greater than 40 and SEX is equal to "M" or AGE greater than 41 and SEX is equal to "F".	DM	SGS-DRG-001-01-S039	16	AGE, SEX	41.0, M
ROMAN-0002	AGE greater than 40 and SEX is equal to "M" or AGE greater than 41 and SEX is equal to "F".	DM	SGS-DRG-001-01-S047	19	AGE, SEX	48.0, F
ROMAN-0002	AGE greater than 40 and SEX is equal to "M" or AGE greater than 41 and SEX is equal to "F".	DM	SGS-DRG-001-01-S048	20	AGE, SEX	44.0, M
ROMAN-0002	AGE greater than 40 and SEX is equal to "M" or AGE greater than 41 and SEX is equal to "F".	DM	SGS-DRG-001-01-S049	21	AGE, SEX	43.0, M
ROMAN-0002	AGE greater than 40 and SEX is equal to "M" or AGE greater than 41 and SEX is equal to "F".	DM	SGS-DRG-001-01-S052	23	AGE, SEX	51.0, M

Use case: define.xml

Validate external data based on a non-SDTM define.xml

- Structure
- Codelist

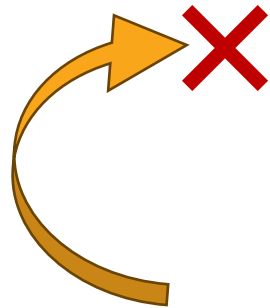
Variable	Label / Description	Type	Role	Length or Display Format	Controlled Terms or ISO Format	Origin / Source / Method / Comment
Trial_id	Study Identifier	text	Identifier	12		Protocol (Source: Sponsor)
Site_id	Site identifier	text	Identifier	2		Assigned (Source: Sponsor)
Screenid	Subject Identifier	text	Identifier	8		Assigned (Source: Sponsor)
Testname	Lab Test Name	text	Topic	6	Labo Test Code "PROTEIN" = "Protein" "MAGNESIUM" = "Magnesium" "CALCIUM" = "Calcium"	Assigned (Source: Sponsor)
Gender	Gender	integer	Identifier	3		Derived (Source: Sponsor) [unresolved: MT.SEQ]
value	Result	text	Topic	6		Assigned (Source: Sponsor)
Units	Units	text	Topic	6		Assigned (Source: Sponsor)

Trial_id	Site_id	Country	Screenid	Testname	Gender	value	Unit
<i>Study Identifier</i>	<i>Site identifier</i>	<i>Country</i>	<i>Subject Identifier</i>	<i>Lab Test Name</i>	<i>Gender</i>	<i>value of the test</i>	<i>Original Units</i>
<i>Char</i>	<i>Char</i>	<i>char</i>	<i>Char</i>	<i>Char</i>	<i>Char</i>	<i>Char</i>	<i>Char</i>
50	50	50	50	50	50	50	50
SGS-DRG-001	L001	BE	SGS-DRG-001-01-S001	PROTEINE	F	4.6	g/dL
SGS-DRG-001	L001	BE	SGS-DRG-001-01-S002	PROTEIN	F	73	U/L
SGS-DRG-001	L001	BE	SGS-DRG-001-01-S003	PROTEINS	M	8	
SGS-DRG-001	L001	BE	SGS-DRG-001-01-S004	CALCIUM	F	9.7	mg/dL
SGS-DRG-001	L001	BE	SGS-DRG-001-01-S005	Magnesium	F	162	mg/dL
SGS-DRG-001	L001	BE	SGS-DRG-001-01-S006	MAGNESIUM	F	23	U/L
SGS-DRG-001	L001	BE	SGS-DRG-001-01-S007	GLUCOSE	M	77	mg/dL
SGS-DRG-001	L001	BE	SGS-DRG-001-01-S008	MAGNESIUM	F	188	U/L
SGS-DRG-001	L001	BE	SGS-DRG-001-01-S009	MAGNESIUM	F	2.1	
SGS-DRG-001	L001	BE	SGS-DRG-001-01-S010	PHOSPHATE	M	4.4	mg/dL
SGS-DRG-001	L003	BE	SGS-DRG-001-01-S011	PROTEIN	F	7.7	g/dL
SGS-DRG-001	L004	BE	SGS-DRG-001-01-S012	ALT	M	16	U/L

Use case: define.xml

Validate external data based on a non-SDTM define.xml

- Structure
- Codelist



Variable	Label / Description	Type	Role	Length or Display Format	Controlled Terms or ISO Format	Origin / Source / Method / Comment
Trial_id	Study Identifier	text	Identifier	12		Protocol (Source: Sponsor)
Site_id	Site identifier	text	Identifier	2		Assigned (Source: Sponsor)
Screenid	Subject Identifier	text	Identifier	8		Assigned (Source: Sponsor)
Testname	Lab Test Name	text	Topic	6	Labo Test Code "PROTEIN" = "Protein" "MAGNESIUM" = "Magnesium" "CALCIUM" = "Calcium"	Assigned (Source: Sponsor)
Gender	Gender	integer	Identifier	3		Derived (Source: Sponsor) [unresolved: MT.SEQ]
value	Result	text	Topic	6		Assigned (Source: Sponsor)
Units	Units	text	Topic	6		Assigned (Source: Sponsor)

Trial_id	Site_id	Country	Screenid	Testname	Gender	value	Unit
Study Identifier	Site identifier	Country	Subject Identifier	Lab Test Name	Gender	value of the test	Original Units
Char	Char	char	Char	Char	Char	Char	Char
50	50	50	50	50	50	50	50
SGS-DRG-001	L001	BE	SGS-DRG-001-01-S001	PROTEINE	F	4.6	g/dL
SGS-DRG-001	L001	BE	SGS-DRG-001-01-S002	PROTEIN	F	73	U/L
SGS-DRG-001	L001	BE	SGS-DRG-001-01-S003	PROTEINS	M	8	
SGS-DRG-001	L001	BE	SGS-DRG-001-01-S004	CALCIUM	F	9.7	mg/dL
SGS-DRG-001	L001	BE	SGS-DRG-001-01-S005	Magnesium	F	162	mg/dL
SGS-DRG-001	L001	BE	SGS-DRG-001-01-S006	MAGNESIUM	F	23	U/L
SGS-DRG-001	L001	BE	SGS-DRG-001-01-S007	GLUCOSE	M	77	mg/dL
SGS-DRG-001	L001	BE	SGS-DRG-001-01-S008	MAGNESIUM	F	188	U/L
SGS-DRG-001	L001	BE	SGS-DRG-001-01-S009	MAGNESIUM	F	2.1	
SGS-DRG-001	L001	BE	SGS-DRG-001-01-S010	PHOSPHATE	M	4.4	mg/dL
SGS-DRG-001	L003	BE	SGS-DRG-001-01-S011	PROTEIN	F	7.7	g/dL
SGS-DRG-001	L004	BE	SGS-DRG-001-01-S012	ALT	M	16	U/L

Use case: define.xml

Validate external data based on a non-SDTM define.xml

- **Structure**

```
Check:
  all:
    - name: variable_name
      operator: not_equal_to
      value: define_variable_name

Core:
  Id: "ROM-004"
  Status: Draft
  Version: '1'
  Description: 'Raise an error when the variable name
in the define.xml does not
correspond to the variable name in the dataset'
  Executability: Fully Executable
  Outcome:
    Message: Variable in dataset not available in the
define.xml
  Rule Type: Variable Metadata Check against Define XML
```

Rule Editor

EDIT

TEST

DIFF

- ✓ Validate YAML Syntax
- ✓ Validate YAML against Schema
- ✓ Convert YAML to JSON Rule
- ✓ Load Test Define.xml
- ✓ Load Test Datasets
- ✓ Results Negatives²

Use case: define.xml

Validate external data based on a non-SDTM define.xml

- **Structure**



```
{ 1 item
  "LB" : [ 1 item
    0 : { 5 items
      "executionStatus" : "success"
      "domain" : "LB"
      "variables" : [ 1 item
        0 : "variable_name"
      ]
      "message" : "Variable in dataset not available in the define.xml"
      "errors" : [ 2 items
        0 : { 2 items
          "value" : { 1 item
            "variable_name" : "Country"
          }
          "row" : 3
        }
        1 : { 2 items
          "value" : { 1 item
            "variable_name" : "Unit"
          }
          "row" : 8
        }
      ]
    }
  ]
}
```

Use case: define.xml

Validate external data based on a non-SDTM define.xml

- Structure
- Codelist

Variable	Label / Description	Type	Role	Length or Display Format	Controlled Terms or ISO Format	Origin / Source / Method / Comment
Trial_id	Study Identifier	text	Identifier	12		Protocol (Source: Sponsor)
Site_id	Site identifier	text	Identifier	2		Assigned (Source: Sponsor)
Screenid	Subject Identifier	text	Identifier	8		Assigned (Source: Sponsor)
Testname	Lab Test Name	text	Topic	6	Labo Test Code • "PROTEIN" = "Protein" • "MAGNESIUM" = "Magnesium" • "CALCIUM" = "Calcium"	Assigned (Source: Sponsor)
Gender	Gender	integer	Identifier			Derived (Source: Sponsor) [unresolved: MT.SEQ]
value	Result	text	Topic	6		Assigned (Source: Sponsor)
Units	Units	text	Topic	6		Assigned (Source: Sponsor)

Trial_id	Site_id	Screenid	Testname	Gender	value	Units
<i>Study Identifier</i>	<i>Site identifier</i>	<i>Subject Identifier</i>	<i>Lab Test Name</i>	<i>Gender</i>	<i>value of the test</i>	<i>Original Units</i>
<i>Char</i>	<i>Char</i>	<i>Char</i>	<i>Char</i>	<i>Char</i>	<i>Char</i>	<i>Char</i>
50	50	50	50	50	50	50
SGS-DRG-001	L001	SGS-DRG-001-01-S001	PROTEINE	F	4.6	g/dL
SGS-DRG-001	L001	SGS-DRG-001-01-S002	PROTEIN	F	73	U/L
SGS-DRG-001	L001	SGS-DRG-001-01-S003	PROTEINS	M	8	
SGS-DRG-001	L001	SGS-DRG-001-01-S004	CALCIUM	F	9.7	mg/dL
SGS-DRG-001	L001	SGS-DRG-001-01-S005	Magnesium	F	162	mg/dL
SGS-DRG-001	L001	SGS-DRG-001-01-S006	MAGNESIUM	F	23	U/L
SGS-DRG-001	L001	SGS-DRG-001-01-S007	GLUCOSE	M	77	mg/dL
SGS-DRG-001	L001	SGS-DRG-001-01-S008	MAGNESIUM	F	188	U/L
SGS-DRG-001	L001	SGS-DRG-001-01-S009	MAGNESIUM	F	2.1	
SGS-DRG-001	L001	SGS-DRG-001-01-S010	PHOSPHATE	M	4.4	mg/dL
SGS-DRG-001	L003	SGS-DRG-001-01-S011	PROTEIN	F	7.7	g/dL
SGS-DRG-001	L004	SGS-DRG-001-01-S012	ALT	M	16	U/L

Use case: define.xml

Validate external data based on a non-SDTM define.xml

- **Codelist**

Check:

all:

- name: define_variable_ccode
operator: non_empty
- name: variable_value
operator: non_empty
- name: define_variable_has_codelist
operator: equal_to
value: true
- name: variable_value
operator: is_not_contained_by
value: define_variable_codelist_coded_values

Rule Editor

EDIT

TEST

DIFF



Validate YAML Syntax



Validate YAML against Schema



Convert YAML to JSON Rule



Load Test Define.xml



Load Test Datasets



Results

Negatives

6

Use case: define.xml

Validate external data based on a non-SDTM define.xml

- **Codelist**



```
"message" : "variable value not present in codelist in the define.xml"
  ▼ "errors" : [ 6 items
    ▼ 0 : { 2 items
      ▼ "value" : { 4 items
        "variable_value" : "PROTEINE"
        ▼ "define_variable_codelist_coded_values" : [ 3 items
          0 : "PROTEIN"
          1 : "MAGNESIUM"
          2 : "CALCIUM"
        ]
        "define_variable_ccode" : "C99999"
        "define_variable_has_codelist" : true
      }
      "row" : 37
    }
    ▼ 1 : { 2 items
```

Use case: define.xml

Validate external data based on a non-SDTM define.xml

- **Structure** ✓
- **Codelist** ✓
- Data types, value length, label, value lists, ...
- Rules are neither trial, vendor, transfer nor standard specific and could therefore potentially be used for any vendor transfer
- Enormous potential
 - Data transfer agreement between sponsor/CRO/Vendor → define.xml
 - Automate validation of the transfers
 - Annotate vendor define.xml to SDTM
 - Automate the SDTM creation

Agenda



- 01 About CDISC Open Rules
- 02 Local custom rules
- 03 Who can create rules and why should we
- 04 Creating custom CDISC Open Rules
- 05 Use cases
- 06 Suggestions – AI integration**
- 07 Key takeaways
- 08 Q&A

Suggestions

- Good rule governance
 - User / roles
 - Categorisation
- Follow-up on the output
 - Producing excel/JSON
- AI integration in the editor

Hello there! My name is COR-E, which rule do you want me to create today?

Hi COR-E, can you please create a rule that checks if a subject has taken more than 250 mg of Paracetamol?

Certainly! Here's the requested rule. Do you want me to create test data for it?

Yes, that would be very helpful

Sure, no problem, here you go!

testData.xlsx

All 7 rules loaded.

Authorities:

- Organization: CDISC

Standards:

- Name: SDTMIG

Version:

Referenc

- Origin: SDTMIG Conformance Rules

Rule Identifier:

Id: CGXXXX

Criteria:

Type: Failure

Version: 2.0

Core:

Id: CORE-000001

Version: 1

Status: Draft

Description: >

This rule checks that no subject has taken more than

Rule Type: Record Data

Scope:

Classes:

Include:

- INTERVENTIONS

Domains:





CDISC CORE Rule Generator

By Roman Radelicki 👤

Generates CDISC CORE rules in YAML code format.

Generate a CDISC
CORE rule for LB
domain and creat...

Create a YAML rule
for the DM domain
and provide test...

Generate test data
for a new rule in
the AE domain.

Provide YAML
code and test data
for a rule in the VS...



Message CDISC CORE Rule Generator



Agenda



- 01 About CDISC Open Rules
- 02 Local custom rules
- 03 Who can create rules and why should we
- 04 Creating custom CDISC Open Rules
- 05 Use cases
- 06 Suggestions – AI integration
- 07 Key takeaways**
- 08 Q&A

Key takeaways

- Custom local rules are very much possible!
- Interchangeability of rules
- Considerations for future success

Questions?



A hand holding a smartphone with a blurred background of city lights.

Contact us

SGS Health Science

Europe +32 15 27 32 45

Americas + 1 877 677 2667

clinicalresearch@sgs.com
www.sgs.com/healthscience

When you need to be sure