



2024 CDISC + TMF  
EUROPE INTERCHANGE

BERLIN

24-25 APRIL: CONFERENCE & EXPO | 22, 23, 26 APRIL: TRAININGS

## **Our Road to Adopting CORE An Implementation Journey on SDTM Dataset Validation**

Presented by Els Janssens  
Data Management System and Process Manager, SGS Health Sciences



## Meet the Speaker

Els Janssens

**Title:** Data Management System and Process Manager

**Organization:** SGS Health Science

Els Janssens has more than 11 years of experience in the field of Clinical Data Management and CDISC SDTM standards (pharma and CRO). In her current role as Data Management Systems and Process Manager, she is DM expert and point of contact concerning DM systems and processes, with her primary focus on data standards and regulatory requirements. She is also CORE volunteer since September 2021.

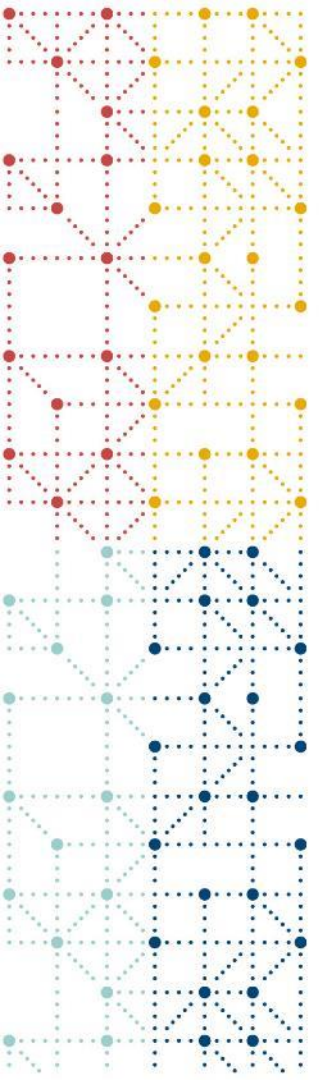
# Disclaimer and Disclosures

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



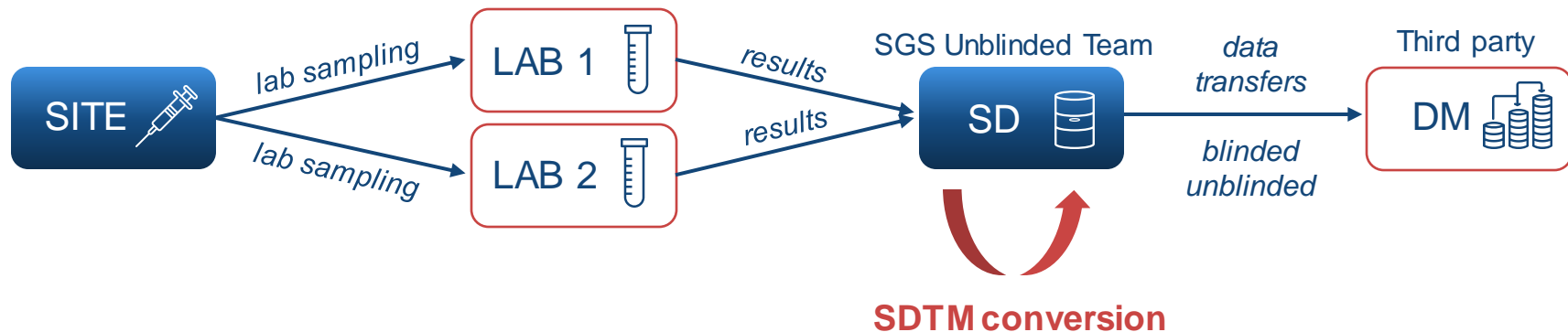
## Agenda

1. Setting the Scene
2. Dataset Validation Meets CORE: Dream Team?
3. Our Journey through Implementation
4. Finetuning our Pathway
5. Conclusions



# 1. Setting the Scene

# Unblinding Data Flow



## Lab samples

- PK samples (blood, stool, urine)
- Blood samples for biomarker analysis or immunogenicity
- ...

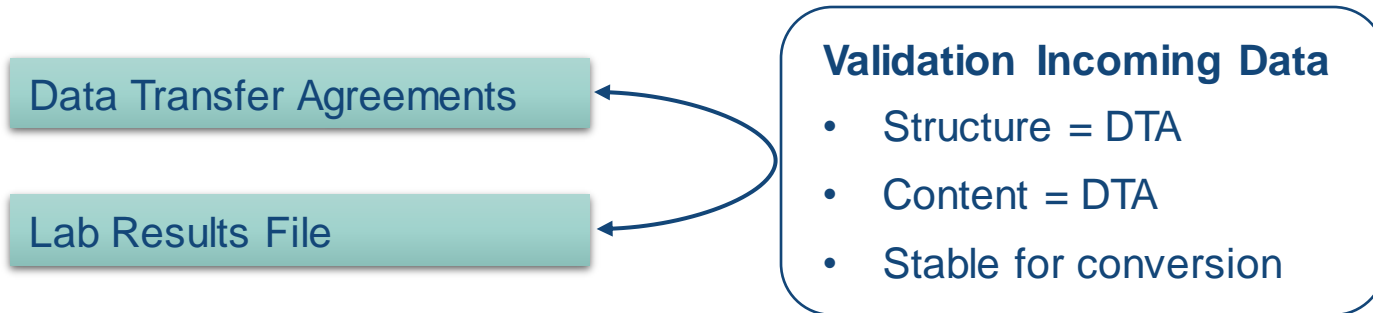
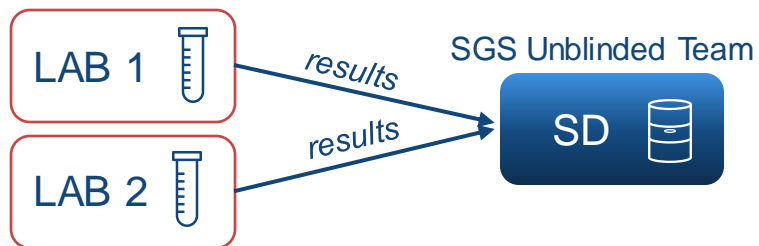
## Datafiles in different structures and formats

- csv
- sas7bdat
- ...

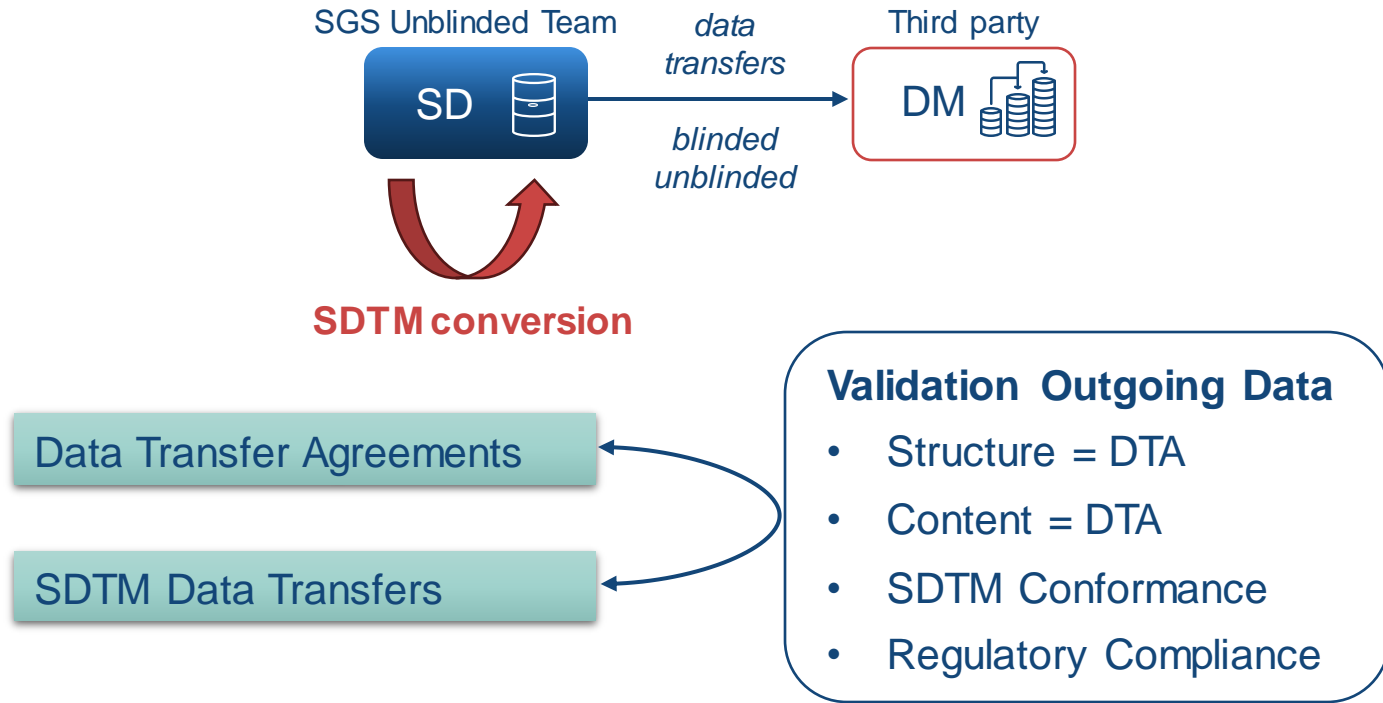
## SDTM datasets

- PC and PP SDTM dataset
- LB and IS SDTM dataset
- ...

# Unblinding Data Flow: Incoming Data Validation



# Unblinding Data Flow: Outgoing Data Validation







## 2. Dataset Validation Meets CORE: Dream Team?

# A New Data Validation Tool

## CURRENT

1. Complex rule output = a list of data = labor intensive

## NEW

1. Rules that generate simple output  
Conformance = no output = no work

# A New Data Validation Tool

## CURRENT

1. Complex rule output = a list of data = labor intensive
2. Programming knowledge needed for rule creation and a lot of documentation

## NEW

1. Rules that generate simple output  
Conformance = no output = no work
2. Easy creation of rules, low code, efficient workflow

# A New Data Validation Tool

## CURRENT

1. Complex rule output = a list of data = labor intensive
2. Programming knowledge needed for rule creation and a lot of documentation
3. Newly released SDTM or regulatory conformance rules need to be added manually = no automatization

## NEW

1. Rules that generate simple output  
Conformance = no output = no work
2. Easy creation of rules, low code, efficient workflow
3. Automatic upload of new SDTM or regulatory conformance rules

# A New Data Validation Tool

## CURRENT

1. Complex rule output = a list of data = labor intensive
2. Programming knowledge needed for rule creation and a lot of documentation
3. Newly released SDTM or regulatory conformance rules need to be added manually = no automatization
4. Rules combine different aspects to be validated = challenging

## NEW

1. Rules that generate simple output  
Conformance = no output = no work
2. Easy creation of rules, low code, efficient workflow
3. Automatic upload of new SDTM or regulatory conformance rules
4. Split up rules in validation purpose = clear categories

# A New Data Validation Tool

## CURRENT

1. Complex rule output = a list of data = labor intensive
2. Programming knowledge needed for rule creation and a lot of documentation
3. Newly released SDTM or regulatory conformance rules need to be added manually = no automatization
4. Rules combine different aspects to be validated = challenging
5. Validates only the SDTM converted outgoing datasets, incoming transfer validated manually

## NEW

1. Rules that generate simple output  
Conformance = no output = no work
2. Easy creation of rules, low code, efficient workflow
3. Automatic upload of new SDTM or regulatory conformance rules
4. Split up rules in validation purpose = clear categories
5. Option to validate incoming vendor file and outgoing SDTM converted dataset

# Is CORE our Superhero?

1. Rules with simple output



```
check:
  all:
    - name: ACTARMCD
      operator: empty
    - name: ARMNRS
      operator: empty
Core:
  Id: CORE-000223
  Status: Published
  Version: '1'
Description: Raise an error when ACTARMCD is empty and ARMNRS is not completed.
Executability: Fully Executable
Outcome:
  Message: ACTARMCD is empty, but ARMNRS is not completed.
Output Variables:
  - ACTARMCD
  - ARMNRS
```

Only output when data is conformant with check  
= SDTM non-conformance

# Is CORE our Superhero?

1. Rules with simple output



2. Easy creation of new rules

CORE Rule Editor available in the SGS environment

Creator	Standards	Orgs	Core ID	Status
<input type="text" value="Janssens, Els"/> ×	<input type="text" value="Search..."/>	<input type="text" value="Search..."/>	<input type="text" value="Search..."/>	<input type="text" value="Search..."/>
Janssens, Els (Mechelen)	SDTMIG	CDISC	SGS-0003	Draft
Janssens, Els (Mechelen)	SDTMIG	CDISC	SGS-0002	Draft



# Is CORE our Superhero?

1. Rules with simple output



2. Easy creation of new rules



## 1. write

```
check:
  all:
    - name: ECOCCUR
      operator: not_equal_to
      value: N
      value_is_literal: true
    - name: ECSTAT
      operator: empty
    - name: ECDOSTXT
      operator: empty
    - name: ECDOSE
      operator: less_than_or_equal_to
      value: 0
```

## 3. check

```
"message" : "ECOCCUR is not 'N', ECSTAT and ECDOSTXT are both empty but
"errors" : [ 4 items
  0 : { 4 items
    "value" : { 5 items
      "ECSTAT" : ""
      "ECSTDTC" : "2012-12-28"
      "ECOCCUR" : "Y"
      "ECDOSE" : -1
      "ECDOSTXT" : ""
```

## 2. test

EDIT TEST DIFF

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

TEST DATASETS FILE...

Filename: unit-test-coreid-CG0100-negative 1.xlsx

✓ Results Negatives 4

## 4. publish

+ [Icons] Search YAML...

Rule Ids Publish Rule Standards

# Is CORE our Superhero?

## CORE Project Objectives on CORE | CDISC

The CORE Project objectives are to:

- Ensure each standard has a set of unambiguous, executable Conformance Rules
- Ensure consistency across Conformance Rule implementations
- Expedite the availability of executable Conformance Rules for new Foundational Standards
- Create executable Conformance Rules vetted by the CDISC standards development teams
- Create a Reference Implementation of an open-source engine that executes the Rules
- Release the open-source engine under the CDISC Open-Source Alliance (COSA)

1. Rules with simple output



2. Easy creation of new rules



3. SDTM + regulatory conformance rules



FDA collaboration

**CDISC is Proud to Announce a Research Collaboration to Incorporate FDA Business Rules into CDISC's Open Rules Engine (CORE)**

Austin, TX - January 16, 2024 - CDISC is proud to announce a research collaboration with the U.S. Food and Drug Administration's Office of Translational Sciences in the Center for Drug Evaluation and Research and Office of Regulatory Operations in the Center for Biologics Evaluation and Research to incorporate FDA Business Rules into CDISC's Open Rules Engine (CORE).

# Is CORE our Superhero?



*SDTM Conformance*

*Regulatory Compliance*

*Data Content*

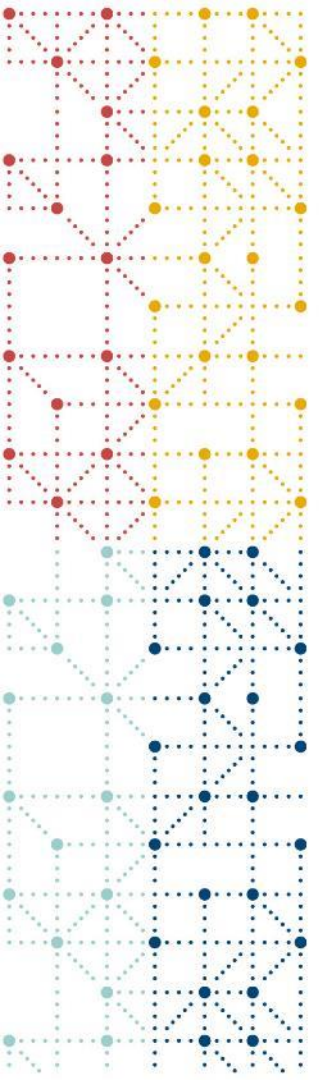
*DTA Compliance*

# Is CORE our Superhero?



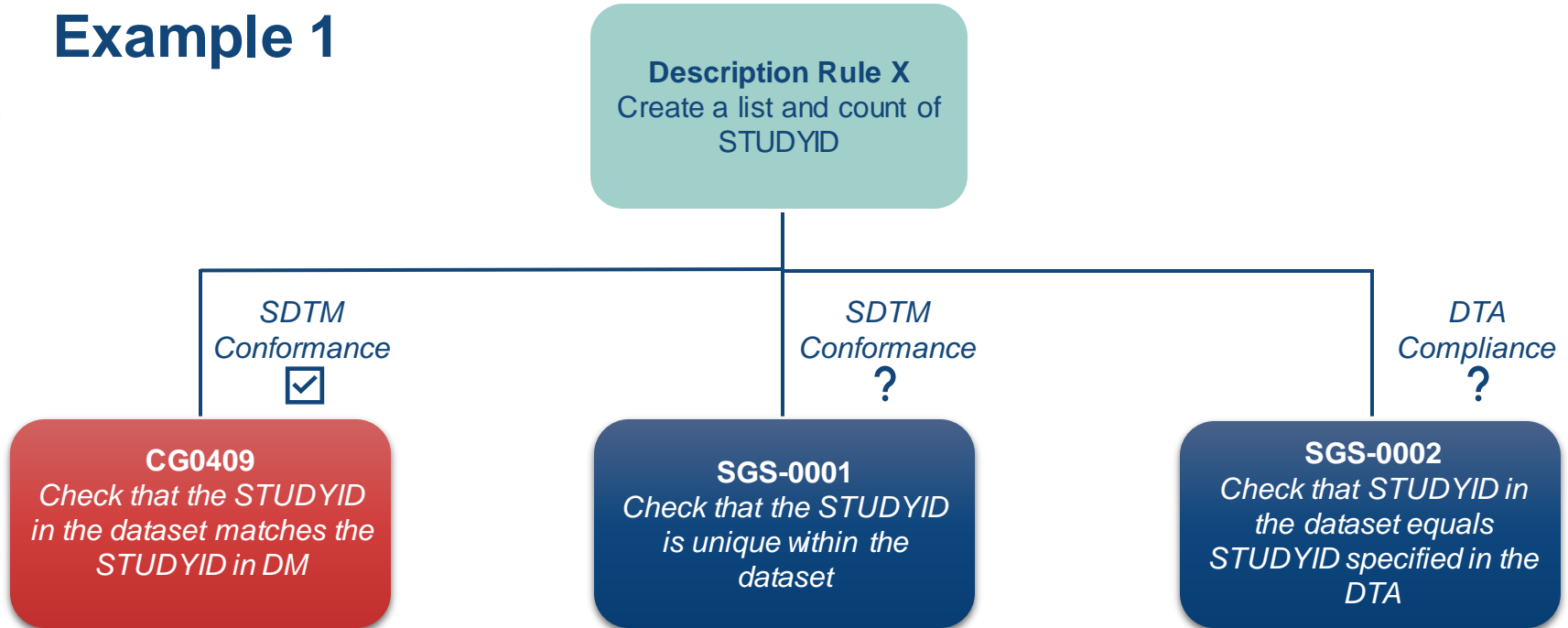
*SDTM Data Structure* ✓

*Other Data Structures* ?



### 3. Our Implementation Journey so Far

# Example 1



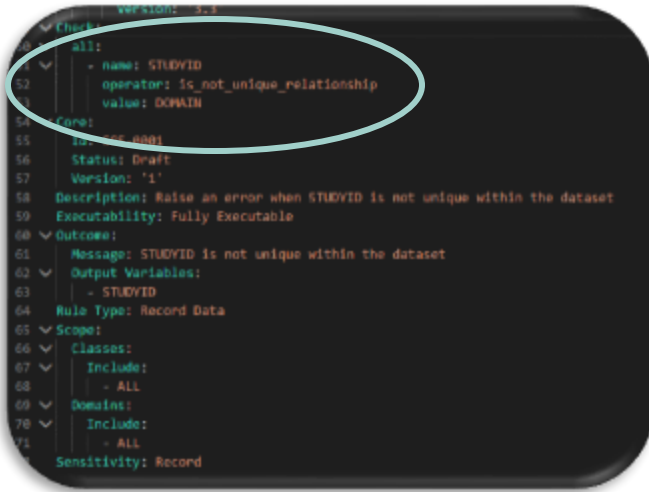
## 4.2.6 Grouping Variables and Categorization

All records with the same STUDYID  
value are a group of records that  
describe that study.

# Example 1: SDTM Data Structure and Conformance

- **SGS-0001**: Check that the STUDYID is unique within the dataset

## 1. write

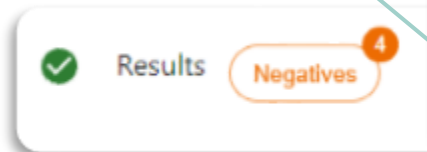


```
version: 3.0
check:
  all:
    - name: STUDYID
      operator: is_not_unique_relationship
      value: DOMAIN
  core:
    - name: SGS-0001
      status: Draft
      version: '1'
      description: Raise an error when STUDYID is not unique within the dataset
      executability: Fully Executable
      outcome:
        message: STUDYID is not unique within the dataset
      outputVariables:
        - STUDYID
      ruleType: Record Data
      scope:
        classes:
          include:
            - ALL
        domains:
          include:
            - ALL
      sensitivity: Record
```

## 2. test

STUDYID	DOMAIN	USUBJID	ISSEQ	ISTESTCD
Study Identifier	Domain Abbreviation	Unique Subject Identifier	Sequence Number	Immunogenicity Test/Exam Short Name
Char	Char	Char		
50	50	50		
CDISCPIL01	IS	CDISCO01		
CDISCPIL01	IS	CDISCO01		
CDISCPIL02	IS	CDISCO01		
CDISCPIL04	IS	CDISCO01		

## 3. check



Results Negatives 4



```
{
  "executionStatus": "success",
  "domain": "IS",
  "variables": [
    {
      "STUDYID": "CDISCPIL01"
    }
  ],
  "message": "STUDYID is not unique within the dataset",
  "errors": [
    {
      "value": [
        {
          "row": 1,
          "USUBJID": "CDISCO01",
          "SEQ": 1
        }
      ]
    },
    {
      "value": [
        {
          "row": 1,
          "USUBJID": "CDISCO01",
          "SEQ": 1
        }
      ]
    }
  ]
}
```

# Example 1: SDTM Data Structure and DTA Compliance

- **SGS-0002**: Check that *STUDYID* in the dataset equals *STUDYID* specified in the DTA

## 1. write

```
all:
  - name: STUDYID
    operator: is_not_contained_by
    value: $dta_studyid

Core:
  Id: SGS-0002
  Status: Draft
  Version: '1'
  Description: Raise an error when STUDYID is not equal to DTA.STUDYID
  Executability: Fully Executable
  Operations:
    - domain: DTA
      id: $dta_studyid
      name: STUDYID
      operator: distinct
  Message: STUDYID is not equal to STUDYID as specified in the DTA
  Output Variables:
    - STUDYID
  Rule Type: Record Data
```

## 2. test

Filename	Label
dta.xpt	Data Transfer Agreements
lb.xpt	Laboratory Test Results

STUDYID	DOMAIN	USUBJID
Study Identifier	Domain Abbreviation	Unique Subject Identifier
Char	Char	Char
12	2	8
CDISCPIL0T01	LB	CDISC001
CDISCPIL0T02	LB	CDISC001
CDISCPIL0T03	LB	CDISC001
CDISCPIL0T04	LB	CDISC001

STUDYID	DOMAIN	USUBJID	LBREFID
Study Identifier	Domain Abbreviation	Unique Subject Identifier	Sequence Number
Char	Char	Char	Num
12	2	8	8
CDISCPIL0T01	LB		



# Example 1: SDTM Data Structure and DTA Compliance

- **SGS-0002**: Check that STUDYID in the dataset equals STUDYID specified in the DTA

## 1. re-write

```
all:
  - name: STUDYID
    operator: is_not_contained_by
    value: $dta_studyid

core:
  id: SGS-0002
  Status: Draft
  Version: '1'
  Description: Raise an error when STUDYID is not equal to DTA.STUDYID
  Executability: Fully Executable
  Operations:
    - domain: DT
      id: $dta_studyid
      name: STUDYID
      operator: distinct
  Outcomes:
    Message: STUDYID is not equal to STUDYID as specified in the DTA
  Output Variables:
    - STUDYID
```

## 2. test

## 3. check

Filename	Label
dt.xpt	Data Transfer Agreements

```
"DT" : [ 1 item
  0 : { 5 items
    "executionStatus" : "success"
    "domain" : "DT"
    "variables" : [ ] 0 items
    "message" : null
    "errors" : [ ] 0 items
  }
  "LB" : [ 1 item
    "message" : "STUDYID is not equal to STUDYID as specified in the DTA"
    "errors" : [ 3 items
      0 : { 4 items
        "value" : { 1 item
          "STUDYID" : "CDISCPILOT02"
        }
      }
    ]
  }
  "row" : 2
```

Results: Positives 1, Negatives 3

success!

# Example 2

**Description Rule Y**  
Create an overview where the STRESC, STRESN and ORRES contents are not the same



*SDTM Conformance*  
☑

**CG0397**  
*Check that --STRESC is completed when --ORRES is completed or DRVFL = 'Y'*

*Regulatory Compliance*  
☑

**FDAB031**  
*Standardized Result in Numeric Format should be populated whenever it is applicable*

*Data Content*  
?

**SGS-0003**  
*Check that --STRESC is equal to --ORRES when --STRESU equals --ORRESU*

*Data Content*  
?

**SGS-0005**  
*Check that --STRESN equals --STRESC when both are populated*

*Data Content*  
?

**SGS-0004**  
*Check that --STRESN is not populated when --STRESC is not numeric*

# Example 2: SDTM Data Structure and Data Content

- **SGS-0003**: Check that --STRESC is equal to --ORRES when --STRESU equals --ORRESU

## 1. write

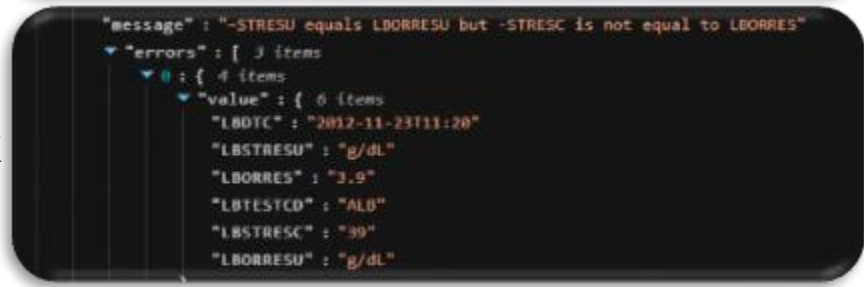


```
rule:
  name: SGS-0003
  operator: not_equal_to
  description: Raise an error when when $STRESU equals --ORRESU but --STRESC is not equal to --ORRES
  executability: Fully Executable
  message: $STRESU equals --ORRESU but --STRESC is not equal to --ORRES
  output_variables:
    - --ORRES
    - --ORRESU
    - --STRESC
    - --ORRESU
  rule_type: Record Data
  scope:
    classes:
      - ALL
    domains:
      - ALL
    (activity): Record
```

## 2. test

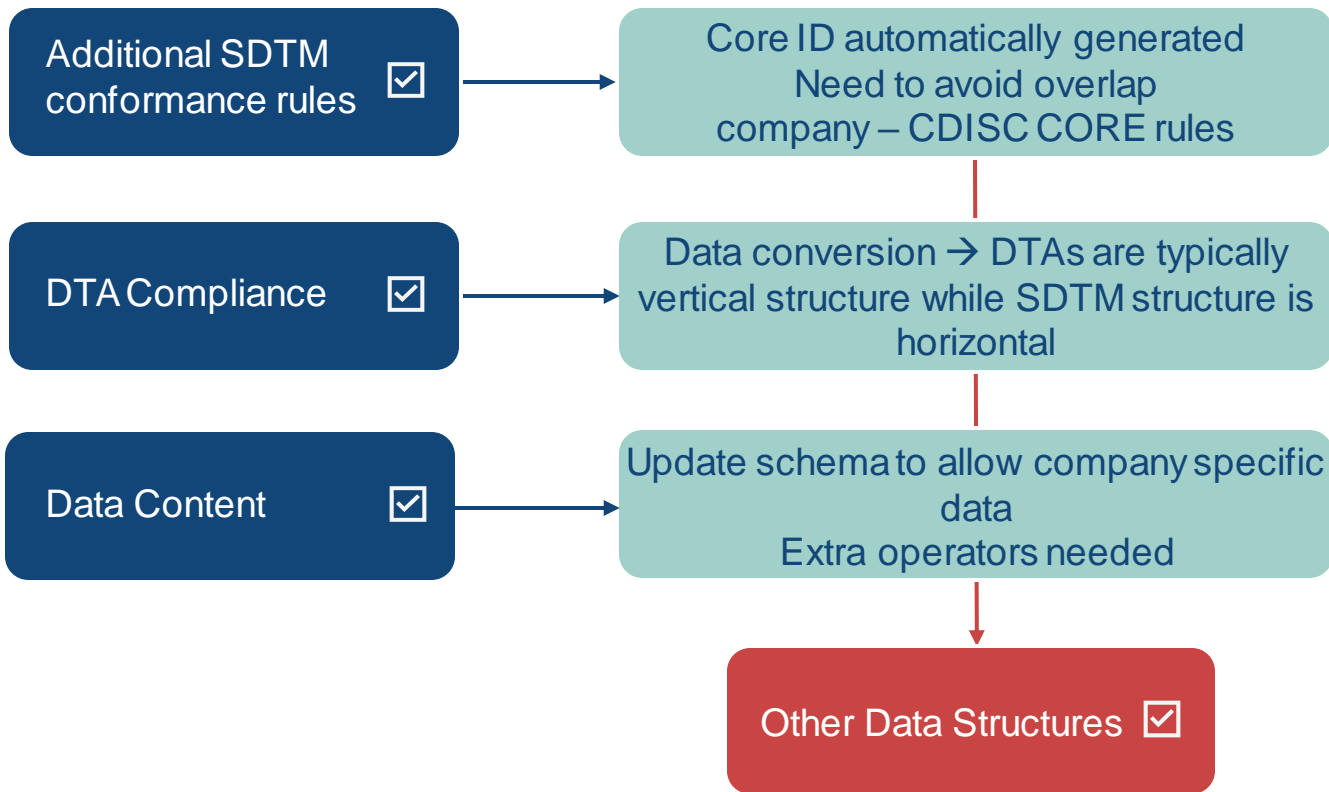
#TESTCD	LBTEST	LBORRES	LBORRESU	LBSTRESC	LBSTRESU	LBOTC
Lab Test or Examination Short Name	Lab Test or Examination Name	Result or Finding in Original	Original Units	Character Result/Finding in Std	Standard Units	Date/Time of Specimen Collection
Char	Char	Char	Char	Char	Char	Char
7	39	6	7	8	7	19
ALB	Albumin	3.9	g/dL	39	g/dL	2012-11-23T11:20
ALP	Alkaline Phosphatase	93	U/L	93	U/L	2012-11-23T11:20
AST	Aspartate Aminotransferase	28	U/L	28	U/L	2012-11-23T11:20

## 3. check



```
{
  "message": "-STRESU equals LBORRESU but -STRESC is not equal to LBORRES"
  "errors": [
    {
      "value": {
        "LBOTC": "2012-11-23T11:20"
        "LBSTRESU": "g/dL"
        "LBORRES": "3.9"
        "LBTESTCD": "ALB"
        "LBSTRESC": "39"
        "LBORRESU": "g/dL"
      }
    }
  ]
}
```

# Insights so far



Core ID	Status
Q Search	Q Search
CORE-000003	Published

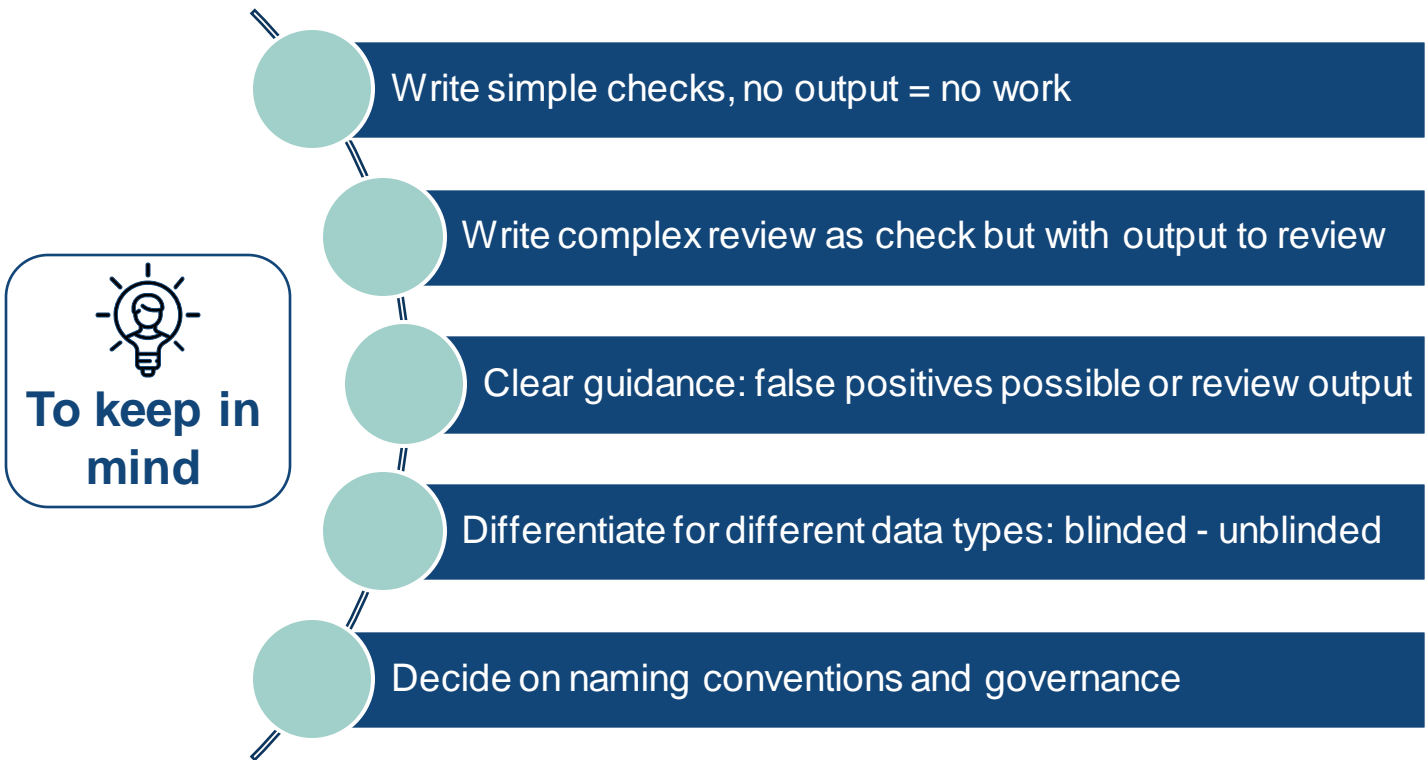
```
Base: SDTMG
References:
- Citations:
  - Cited Guidance: All records
    that describe that study
  Document: 10
  Item: 2.9.3
  Section: "4.2.6"
Origin:
Rule Identifier:
Id: C09999
Version: 3
Version:
Version: "3.3"
```

- Operator.md
- Organization\_CDISC.json
- Organization\_FDA.json
- Organization\_INTERNAL.json
- Organization\_PMDA.json
- Rule\_Type.json
- Rule\_Type.md



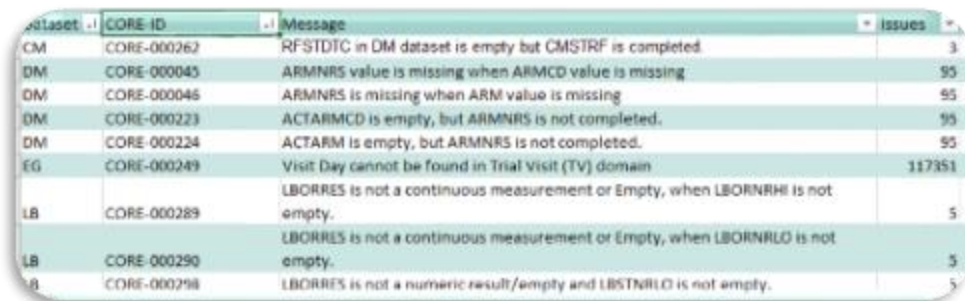
## 4. Finetuning our Pathway

# Rule Creation



# The ideal validation dashboard

- Rule Engine generates output in Excel



Dataset	CORE ID	Message	Issues
CM	CORE-000262	RFSTDTC in DM dataset is empty but CMSTRF is completed.	3
DM	CORE-000043	ARMNRS value is missing when ARMCD value is missing	95
DM	CORE-000046	ARMNRS is missing when ARM value is missing	95
DM	CORE-000223	ACTARMCD is empty, but ARMNRS is not completed.	95
DM	CORE-000224	ACTARM is empty, but ARMNRS is not completed.	95
EG	CORE-000249	Visit Day cannot be found in Trial Visit (TV) domain	117351
LB	CORE-000289	LBORRES is not a continuous measurement or Empty, when LBORNRIH is not empty.	5
LB	CORE-000290	LBORRES is not a continuous measurement or Empty, when LBORNRILO is not empty.	5
LB	CORE-000298	LBORRES is not a numeric result/empty and LBSTNRLO is not empty.	5

- There is a need to
  - develop a more user friendly platform
  - include specific requirements



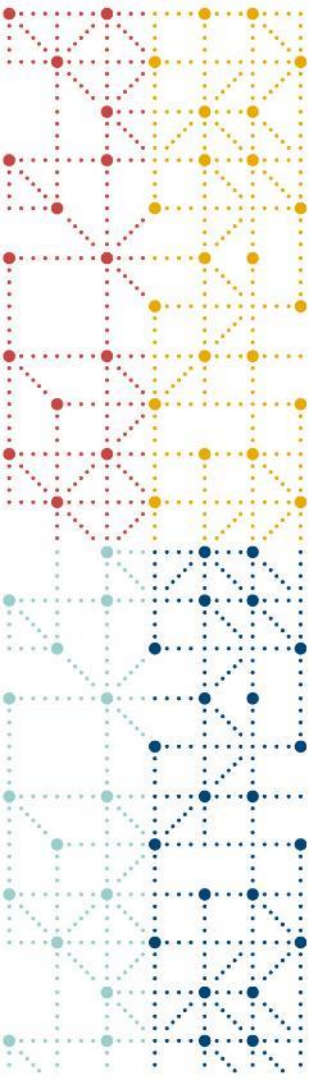


# The ideal validation dashboard

- **Input**
  - Easy upload of datasets for validation
  - Option to validate datasets alone
  - Quick selection of rules
- **Output**
  - Optimal performance = output within minutes
  - Secure environment for validation of unblinding dataset
  - Different environments: test versus production
  - Visually informative and self-explanatory
  - Flag statuses, track changes, add comments
  - Automated queries or reconciliation
  - Statistics







## 5. Conclusions

# Conclusion

- **Yes, we are on the right track!**
  - CORE does meet the requirements for our dataset validation
  - We believe this dream team has great potential
- **To be continued...**
  - Implementation project in progress
  - Perform gap analysis on current rule set and create new rules
  - Implement this not only for unblinding data but for all data
  - Further development of code (editor + engine) → discussions with CDISC needed based on our findings





# SGS team work

Marisa Wyckmans, Chris Fransen, Hannes De Bondt, Roman Radelicki

## Thank You!

### Contact us

[clinicalresearch@sgs.com](mailto:clinicalresearch@sgs.com)

EUROPE: +32 15 27 32 45  
AMERICAS: +1 877 677 2667

[www.sgs.com/cro](http://www.sgs.com/cro)

[health-nutrition@sgs.com](mailto:health-nutrition@sgs.com)  
[www.sgs.com/healthnutrition](http://www.sgs.com/healthnutrition)

**cdisc**

**SGS**