

CDISC CORE

31st German CDISC User Network Meeting, Munich, 2024-03-14

Jozef Aerts
XML4Pharma

What is CDISC CORE?

- CORE = CDISC Open Rules Engine
- Meant to replace the "jungle" of **validation rules** and implementations
- Open source, vendor-neutral, extensible
- All implementations should give the same results

CDISC CORE and FDA

- FDA has asked CDISC to implement the "FDA Business Rules" into CORE
<https://www.cdisc.org/news/cdisc-proud-announce-research-collaboration-incorporate-fda-business-rules-cdiscs-open-rules>
- It is expected that FDA will replace P21 by CORE this year or next year

Who is developing the CORE rules?

- CORE rules are developed starting from the existing CDISC validation rules (mostly as Excel) by a large number of volunteers (>100)
- Engine programming in Python is done by 2+ specialists at CDISC
- Everything is published as "open source" on Github:
<https://github.com/cdisc-org/cdisc-rules-engine/>

Rule implementation made easy by YAML

YAML Rule editor used by the volunteers

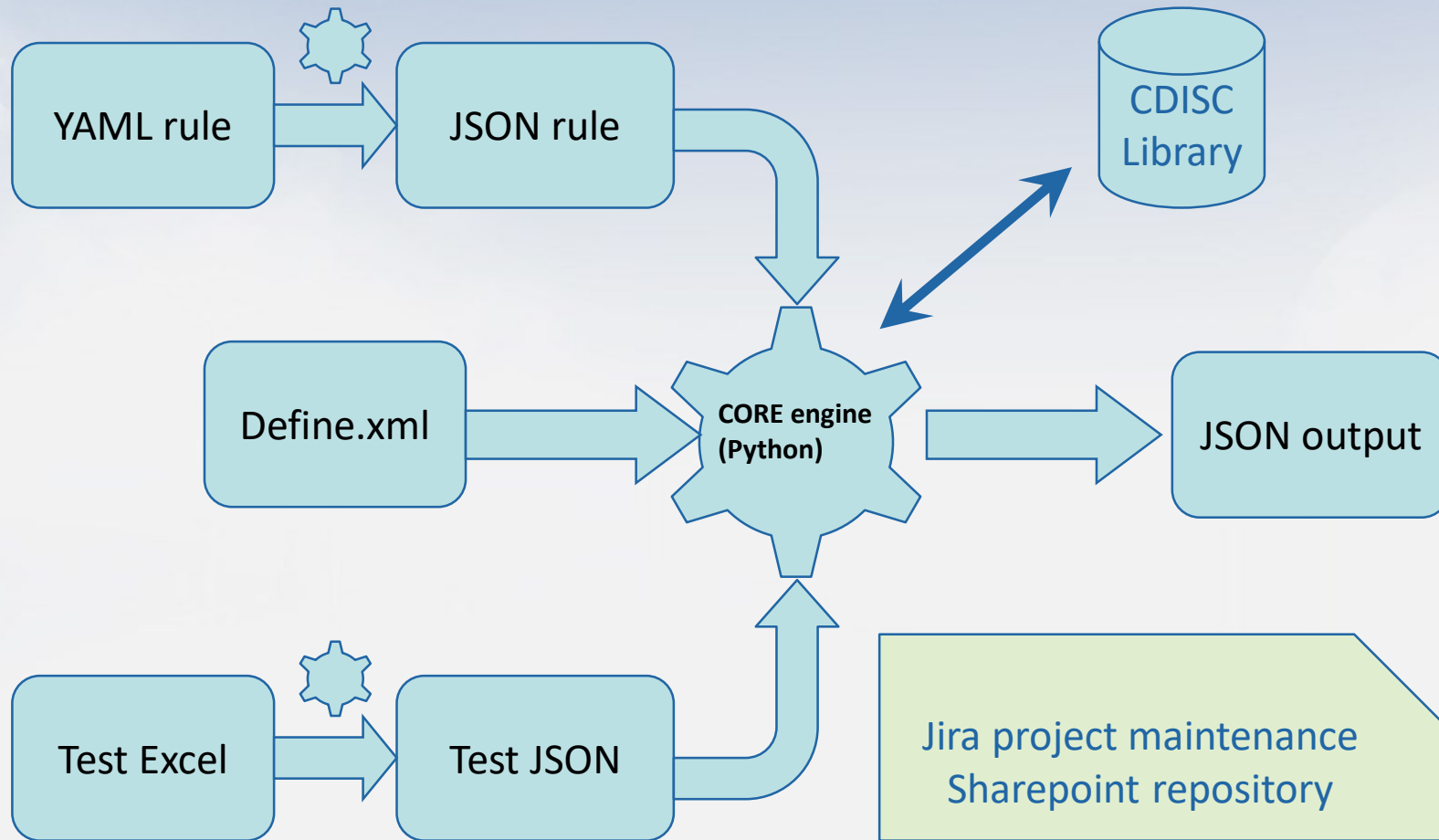
```
EDIT TEST DIFF
1 Authorities:
2 - Organization: FDA
3 Standards:
4 > - Name: SDTMIG ...
15 > - Name: SDTMIG ...
26 > - Name: SDTMIG ...
37 - Name: SDTMIG
38 References:
39 - Citations:
40 - Cited Guidance: FDAB038
41 Document: FDA
42 Origin: FDA Validator Rules
43 Rule Identifier:
44 Id: SD2271
45 Version: '1'
46 Version: '1.6'
47 Version: '3.3'
48 > - Name: SENDIG-AR ...
59 Check:
60 all:
61 - name: --STDTC
62 operator: not_exists
63 - name: --STDY
64 operator: exists
65 Core:
66 Id: FDA.SDTMIG.SD2271
67 Status: Draft
68 Version: '1'
69 Description: Start Study Date/Time (--STDTC) variable should be included into
70 dataset, when Study Day of Start (--STDY) variable is present.
71 Executability: Fully Executable
72 Outcome:
73 Message: Missing --STDTC variable, when --STDY variable is present
74 Rule Type: Record Data
75 Scope:
76 Classes:
77 Include:
78 - ALL
79 Domains:
80 Include:
81 - ALL
82 Sensitivity: Dataset
```

Test files as Excel

	A	B	C	D	E	F	G
1	STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	QNAM	QLABEL
2	Study	Related Domain	Unique Subject	Identifying	Identifying	Qualifier Variable Name	Qualifier Variable Label
3	Char	Char	Char	Char	Char	Char	Char
4	10	2	14	6	4	11	49
5	CDISC-TEST	LB	CDISC-TEST-001	LBSEQ	1	BIOSIG1TOOLONG	Biological Significance 1
6	CDISC-TEST	LB	CDISC-TEST-001	LBSEQ	10	BIOSIG2	Biological Significance 2
7	CDISC-TEST	LB	CDISC-TEST-001	LBSEQ	11	BIOSIG3	Biological Significance 3
8	CDISC-TEST	LB	CDISC-TEST-001	LBSEQ	12	BIOSIG4	Biological Significance 4
9	CDISC-TEST	LB	CDISC-TEST-001	LBSEQ	13	BIOSIG5	Biological Significance 5
10	CDISC-TEST	LB	CDISC-TEST-002	LBSEQ	1	BIOSIG1	Biological Significance 1
11	CDISC-TEST	LB	CDISC-TEST-002	LBSEQ	2	BIOSIG2	Biological Significance 2
12	CDISC-TEST	LB	CDISC-TEST-002	LBSEQ	3	BIOSIG3	Biological Significance 3
13	CDISC-TEST	LB	CDISC-TEST-002	LBSEQ	4	BIOSIG4	Biological Significance 4
14	CDISC-TEST	LB	CDISC-TEST-002	LBSEQ	5	BIOSIG5	Biological Significance 5

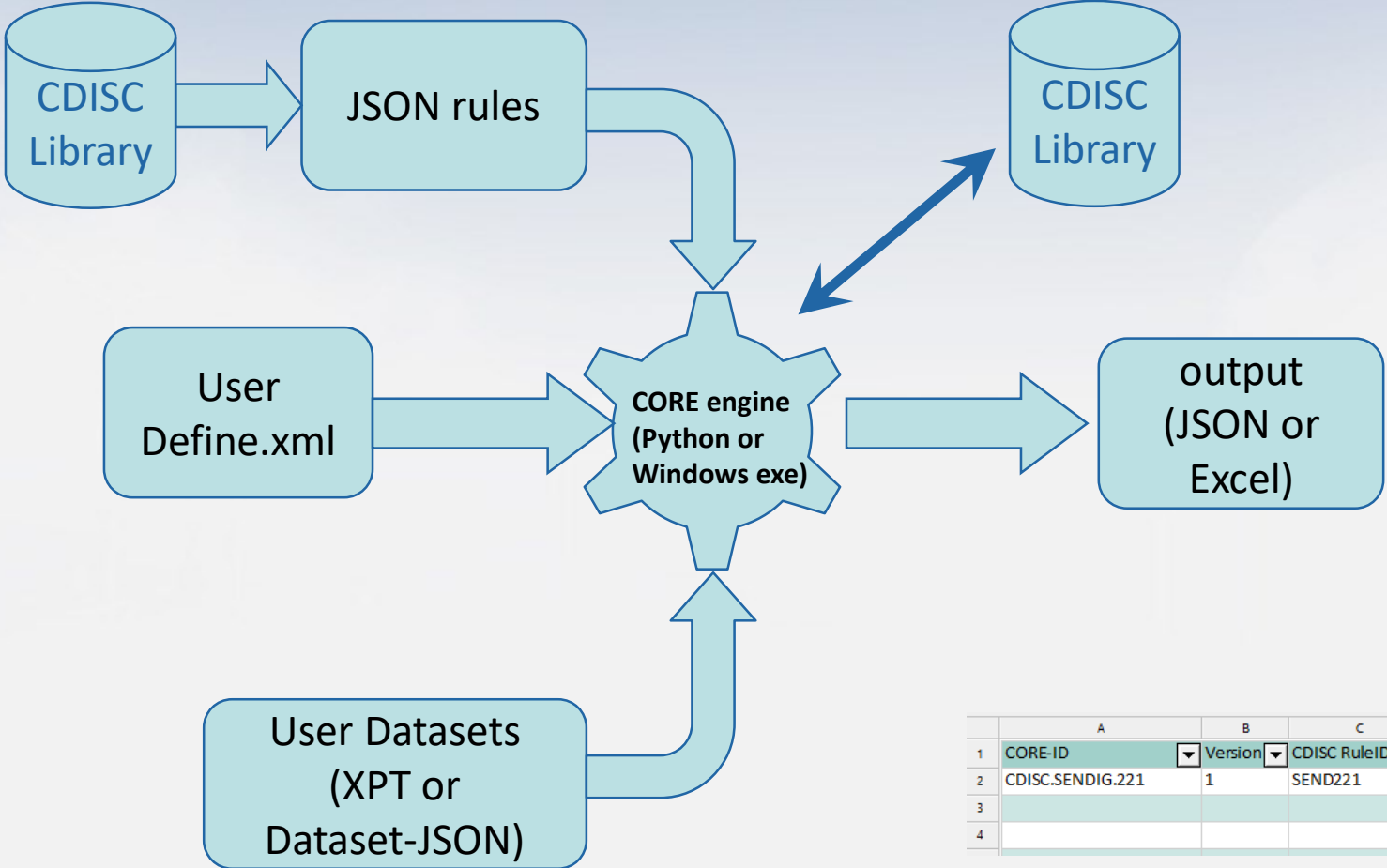
Jozef A

How does that work? Development



```
Results
[ 1 item
  0: { 5 items
    "executionStatus": "success"
    "domain": "CO"
    "variables": [ 2 items
      0: "COSEQ"
      1: "POOLID"
    ]
    "message": "The value of COSEQ is not unique within POOLID"
    "errors": [ 2 items
      0: { 3 items
        "value": { 2 items
          "POOLID": "POOL1"
          "COSEQ": 1
        }
        "row": 5
        "SEQ": 1
      }
      1: { 3 items
        "value": { 2 items
          "POOLID": "POOL1"
          "COSEQ": 1
        }
        "row": 7
        "SEQ": 1
      }
    ]
  }
]
```

How does that work? Production



```

1  {
2  "DM": [
3  {
4    "executionStatus": "success",
5    "domain": "DM",
6    "variables": [
7      "SETCD"
8    ],
9    "message": "SETCD value is greater than 8",
10   "errors": [
11   {
12     "value": {
13       "SETCD": "SETTOOLONG"
14     },
15     "row": 4,
16     "uSubjId": "015246-099-0000-00003"
17   }
18 ]
19 }
20 ],
21 "TX": [
22 {
23   "executionStatus": "success",
24   "domain": "TX",
25   "variables": [
26     "SETCD"
27   ],
28   "message": "SETCD value is greater than 8",
29   "errors": [
30   {
31     "value": {
32       "SETCD": "SETMUCHTOOLONG"
33     },
34     "row": 4,
35     "seq": 4
36   }
37 ]
38 }
39 ]
40 }

```

	A	B	C	D	E	F	G
1	CORE-ID	Version	CDISC RuleID			Message	Status
2	CDISC.SENDIG.221	1	SEND221			TAETORD must be an integer number	SUCCESS
3							
4							

How do I use it?

- "Quick start" on Github
- Essentially using CLI (Command Line Interface)
- Commands are valid for both Python as Windows-exe implementation
- Also instructions how to compile (latest version) to Windows exe

Running a validation

From the command line

Clone the repository and run `python core.py --help` to see the full list of commands.

Run `python core.py validate --help` to see the list of validation options.

```
-ca, --cache TEXT           Relative path to cache files containing pre
                             loaded metadata and rules
-ps, --pool-size INTEGER    Number of parallel processes for validation
-d, --data TEXT             Path to directory containing data files
-dp, --dataset-path TEXT    Absolute path to dataset file. Can be specified
-dxp, --define_xml_path TEXT Path to Define-XML
-l, --log-level [info|debug|error|critical|disabled|warn]
                             Sets log level for engine logs, logs are
                             disabled by default
-rt, --report-template TEXT File path of report template to use for
                             excel output
-s, --standard TEXT         CDISC standard to validate against
                             [required]
-v, --version TEXT          Standard version to validate against
                             [required]
-ct, --controlled-terminology-package TEXT
                             Controlled terminology package to validate
                             against, can provide more than one
-o, --output TEXT           Report output file destination
-of, --output-format [JSON|XLSX]
                             Output file format
-rr, --raw-report           Report in a raw format as it is generated by
                             the engine. This flag must be used only with
                             --output-format JSON.
-dv, --define-version TEXT  Define-XML version used for validation
-dxp, --define-xml-path    Path to define-xml file.
```

How do I use it? List of Rules

- list rules:

- list-rules - list rules available in the cache

- list all rules:

```
python core.py list-rules
```

- list rules for standard:

```
python core.py list-rules -s sdtmig -v 3-4
```

- list-rule-sets - lists all standards and versions for which rules are available: `python core.py list-rule-sets`

- update rules
(requires CDISC-Library
API key):

- update-cache - update locally stored cache data (Requires an environment variable - `CDISC_LIBRARY_API_KEY`)

```
python core.py update-cache
```

To obtain an api key, please follow the instructions found here: <https://wiki.cdisc.org/display/LIBSUPRT/Getting+Started%3A+Access+to+CDISC+Library+API+using+API+Key+Authentication>. Please note it can take up to an hour after sign up to have an api key issued

Implementation in SDTM-ETL

Core Rules Exploration

Filter Reset Select all rules shown Clear selection

Rule ID Sensitivity Authorities Standards and Versions Classes Included Classes Excluded
 Domains Included Domains Excluded Rule Description Message References

Include	Rule ID	Authority	Sensitivity	Standards and Versions	Classes Included	Classes Excluded	Domains Included	Domains Excluded	Description	Message	References
<input type="checkbox"/>	CORE-000280	FDA	Dataset	SDTMIG 3.1.2 SDTMIG 3.1.3 SDTMIG 3.2 SDTMIG 3.3	TRIAL DESIGN		TS		'Therapeutic Area' (THERAREA) rec...	Missing THERAR...	FDA Validator RulesSDTMIG: FDA: F...
<input type="checkbox"/>	CORE-000278	FDA	Dataset	SDTMIG 3.1.2 SDTMIG 3.1.3 SDTMIG 3.3	TRIAL DESIGN		TS		'SDTM IG Version' (SDTIGVER) recor...	Missing SDTIGVE...	FDA Validator RulesSDTMIG: FDA: F...

Execute Transformation (XSLT) C

ODM file with clinical data:

D:\SDTM-ETL\TestFiles\ODM1-3-1\

MetaData in separate ODM file

D:\SDTM-ETL\TestFiles\ODM1-3-1\

A	B	C	D	E	F	G	H	
CORE-ID	Message	Executability	Dataset	USUBJID	Record	Sequence	Variable(s)	Value(s)
CORE-000286	AGE is not provided	fully executable	DM	001		1	AGE, BRTHDTC	None, 1957-05-07

View Result SDTM tables

Generate 'NOT DONE' records

Save Result SDTM tables as SAS XPORT files directory:

D:\temp\

Add location of SAS XPORT file

Additionally generate a merge

Messages and error messages:

D:\temp\core-validation-result_20240304T112821_912.xlsx

OK

<input type="checkbox"/>	CORE-000281	FDA	Record	SDTMIG : SDTMIG : SDTMIG : SDTMIG :				ent, Exposure o... -STDTC is after ...	FDA Validator RulesSDTMIG: FDA: F...
<input type="checkbox"/>	CORE-000279	FDA	Dataset	SDTMIG 3.1.2 SDTMIG 3.1.3 SDTMIG 3.2 SDTMIG 3.3	TRIAL DESIGN		TS	'SDTM Version' (SDTMVER) record ...	Missing SDTMVE... FDA Validator RulesSDTMIG: FDA: F...
<input type="checkbox"/>	CORE-000277	FDA	Dataset	SDTMIG 3.1.2 SDTMIG 3.1.3 SDTMIG 3.2 SDTMIG 3.3	TRIAL DESIGN		TS	'Rare Disease Indicator' (RDIND) re...	Missing RDIND Tr... FDA Validator RulesSDTMIG: FDA: F...
<input type="checkbox"/>	CORE-000288	FDA	Record	SDTMIG 3.1.3 SDTMIG 3.2 SDTMIG 3.3 SENDIG 3.0	SPECIAL PURPOSE		DM	Age Range (AGETXT) variable value...	Invalid value for A... FDA Validator RulesSDTMIG: FDA: F...
<input type="checkbox"/>	CORE-000276	FDA	Record	SDTMIG 3.1.3 SDTMIG 3.2 SDTMIG 3.3	TRIAL DESIGN		TS	Value for the Null Flavor (TSVALNF) v...	Value for TSVALN... FDA Validator RulesSDTMIG: FDA: F...

Close

OK

STUDID	DOMAIN	USUBJID	SS.SSSEQ	SS.SSGRPID
STUDYID	DOMAIN	USUBJID	SS.SSSEQ	SS.SSGRPID
STUDYID	DOMAIN	USUBJID	TU.TUSEQ	TU.TUGRPID



Demo time!

Using CORE in the open source
Smart Submission Dataset Viewer