Paper DS07

CDISC Conformance rules and the CDISC Open Rules Engine
Continuing the Road to Adoption

Peter Van Reusel, CDISC, Austin TX, USA
Nick De Donder, CDISC, Austin TX, USA

ABSTRACT

CDISC Conformance Rules are an integral part of the Foundational Standards and serve as the specific guidance to Industry for the correct implementation of the Standards in clinical studies. The overall goal of the CORE Initiative is to provide a governed set of unambiguous and executable Conformance Rules for each Foundational Standard, and to provide an open-source execution engine for the executable Rules which are available from the CDISC Library. This presentation will begin with a brief review of the CDISC CORE program concept --- what it is and why it is important. Progress and status of development of the CDISC Conformance Rules and the CORE Engine will then be covered. The focus then shifts to uptake and implementation to date by the CDISC community including Pharma and Biotech end-users, software vendors offering CORE solutions, and Regulatory Agencies. Finally, the CDISC Conformance Rules governance model and the CORE roadmap will be discussed.

CORE CONCEPT

CDISC Conformance Rules are an integral part of the CDISC Foundational Standards and serve as the specific guidance to Industry for the correct implementation of the Standards in clinical studies. The overall goal of the CORE (CDISC Open Rules Engine) Project is to deliver a governed set of unambiguous and executable Conformance Rules for each Foundational Standard, and to provide a reference implementation of an open-source execution engine for the executable Rules which are retrieved from the CDISC Library.

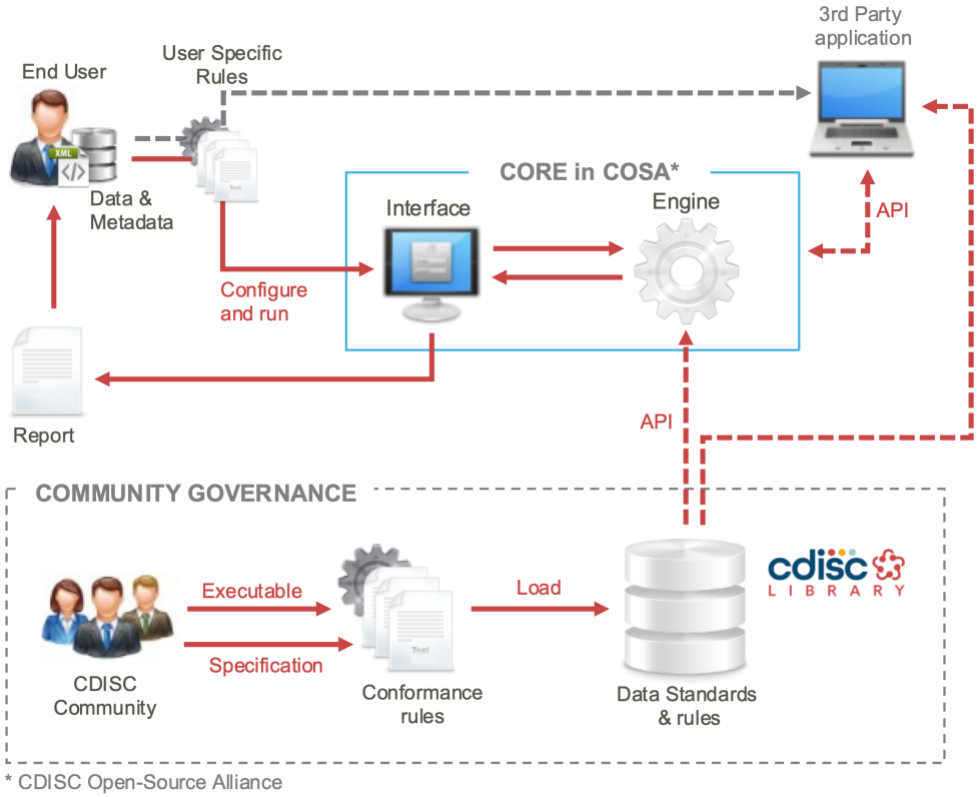
CDISC is conducting the CORE Program for the following reasons:

- 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
- Develop an open-source engine that serves as a Reference Implementation
- Publish the Rules in the CDISC Library and the engine under the CDISC Open-Source Alliance (COSA)

Figure 1 below illustrates the CORE concept. CORE is comprised of both the Conformance Rules (“Rules”) and the Conformance Rules Engine (“Engine”). The Rules are being developed by CDISC per the governed process for all CDISC Standards development, with help from the CDISC Community. There are two expressions of the Rules: (1) the specification (human-readable) and (2) the executable (machine-executable). The Rules are stored in the CDISC Library along with the rest of the CDISC Standards.

The Engine is an open-source software application whose purpose is to execute the Rules against clinical data and return results. The Engine is made available to the CDISC Community in GitHub and users may deploy it in a variety of processing environments including cloud, on-premises desktop, and on-premises server. A user interface (UI) will also be provided in GitHub which may optionally be used to control Engine execution. Users will also deploy their clinical study data within the same overall processing environment. The Engine accesses the Rules from the CDISC Library via a Library API when it executes. Users may also add custom Rules for processing. Finally, an alternate engine may be developed (by an end-user or by a commercial software vendor) to execute the Rules. Such an engine can access the Rules via the Library API and can be validated using a validation package that will be provided by CDISC with the correct results when executing each Rule using the CDISC Engine which is the Reference Implementation.

FIGURE 1: CORE CONCEPT



CONFORMANCE RULES

RULE DEVELOPMENT PROCESS

The Rule development process is one of expressing the human-readable Rule specifications in machine-executable form. This is done by enter each Rule into the CORE Rule Editor using a structured language (YAML). The specific syntax for expressing a Rule using YAML was developed by the CORE team, and the Engine has been designed to ingest and execute the Rules per the YAML instructions. To express each new Rule in the Rule Editor, Rule developers start with a template or copy from an existing similar Rule. The Rule Editor provides hints and completion suggestions and structures the syntax for visual review. There is 1 CORE rule per file which contains the metadata and the logic.

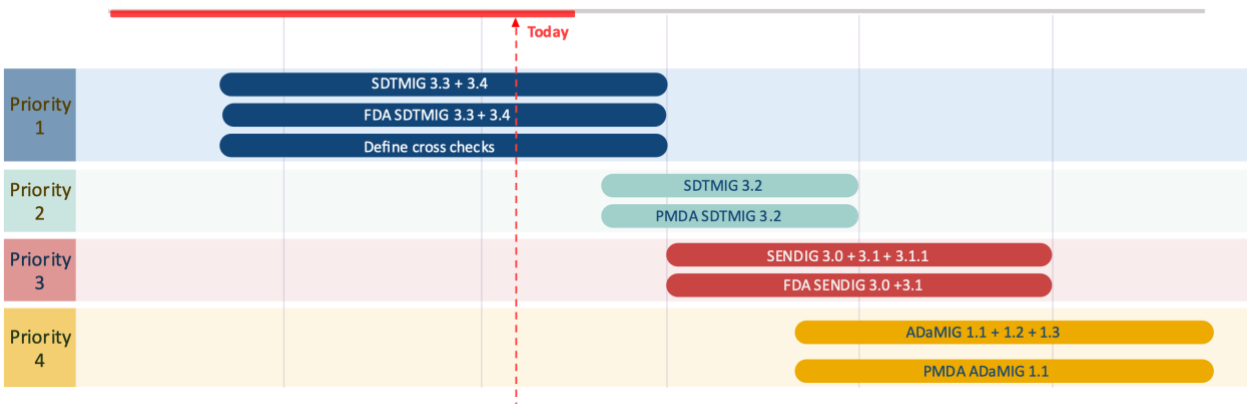
VOLUME AND BREADTH OF CONFORMANCE RULES

As might be expected, there will be many Rules involved to test conformance to the various CDISC Foundational Standards. The Rule set for each Foundational Standards is comprised of two types: those that are executable and those that are not. Looking at volume, there are over 1900 unique rules identified, each referring to information available in the different Implementation Guides. Note that there will be Rules beyond those defined by CDISC for its Foundational Standards. Specifically, the CORE Program will also create executable Rules for these other Rule sets: Regulatory Rules FDA, Regulatory Rules PMDA, and Define-XML Cross-Check.

RULE DEVELPOMENT PRIORITY

The actual timeline for Rule development is challenging to pin down for three primary reasons: (1) the number of executable Rules per Foundational Standard is not yet known, (2) the time needed per Rule varies by the Rule complexity (low, medium, high) and the complexity breakdown is not yet known, and (3) the number of volunteers available for Rule development is not yet known. Nonetheless, making best-estimate assumptions for each of these three dimensions, with the experience had to data, CDISC has set out a priority as outlined in Figure 2 as a target.

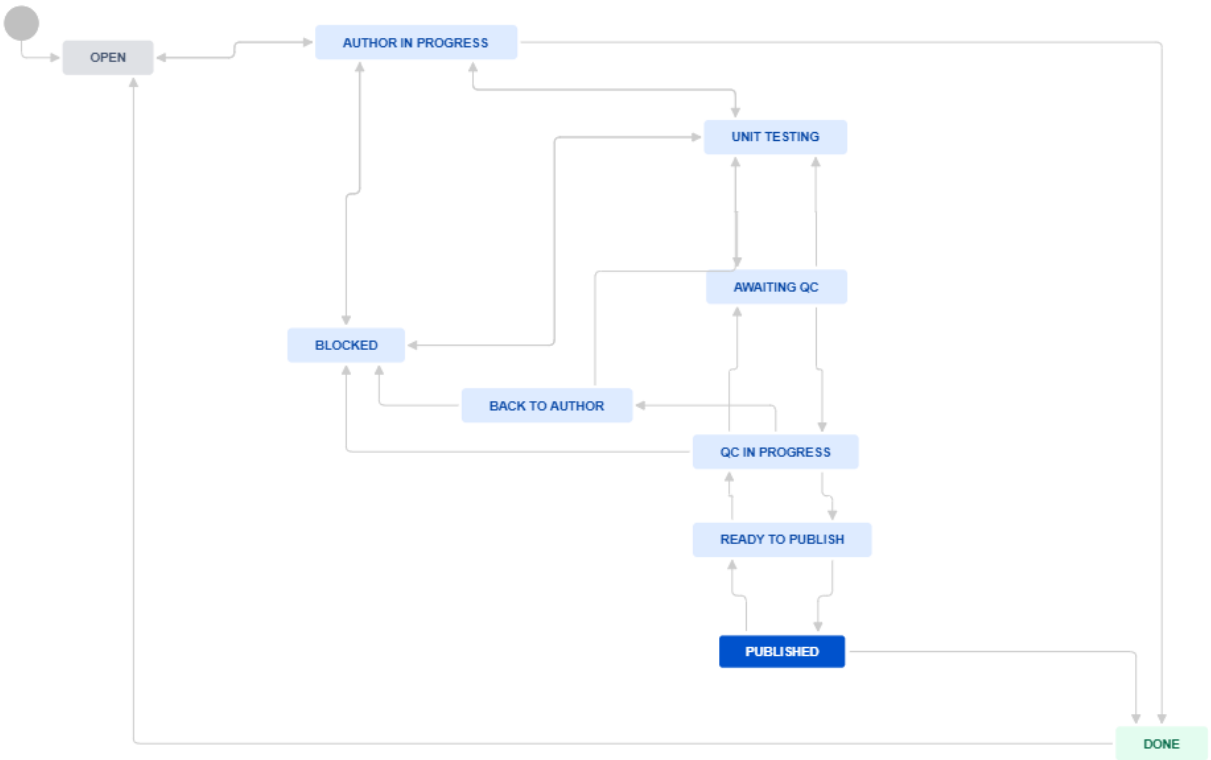
FIGURE 2: RULES DEVELOPMENT PRIORITY



RULE PROCESS

The rule development process is defined as a Jira project. The workflow can be found in Figure 3. Each individual Rule gets written and specified in YAML code in the Rule Editor by a Rule Author. The Rule Author assigns a rule to him/herself and changes the status to “Author in Progress”. Once the rule has been written, it goes into unit testing. As part of unit testing, positive and negative test data gets created to check whether the rule is working as expected. After unit testing a second Rule Author will perform a Quality Control (QC) of the written and confirm whether it is giving the expected outcome. Once confirmed the rule can be published or it can go back to the author to perform additional updates. If the Rule Author notices that the rule cannot be written because additional development of the engine is needed, the rule will be put to “Blocked”. As there are also “non-executable” rule, the status of a rule can be changed to “Done”.

FIGURE 3: RULE DEVELOPMENT WORKFLOW



DEVELOPING CORE RULES IN THE FUTURE

In the future CDISC plans not to generate the rule specifications in an Excel file but directly in the Rule Editor itself. It allows for everything being together, the rule logic, the cited guidance, the rule description, the outcome message, and

the outcome variables. It will however be possible to do an extract of from the Rule Editor so it can be reviewed during internal and public review.

RULES GOVERNANCE MODEL

CORE RULES GOVERNANCE

CDISC aims to establish a CORE governance team which comprises of representatives from CDISC, the Regulatory agencies and the community. The goal of the team is to receive and review specifications that will be loaded in the Central Specifications Log. Different activities will be:

- Govern the specifications content
- Curate the content
- Mark rules for development
- Prioritize rules for development.

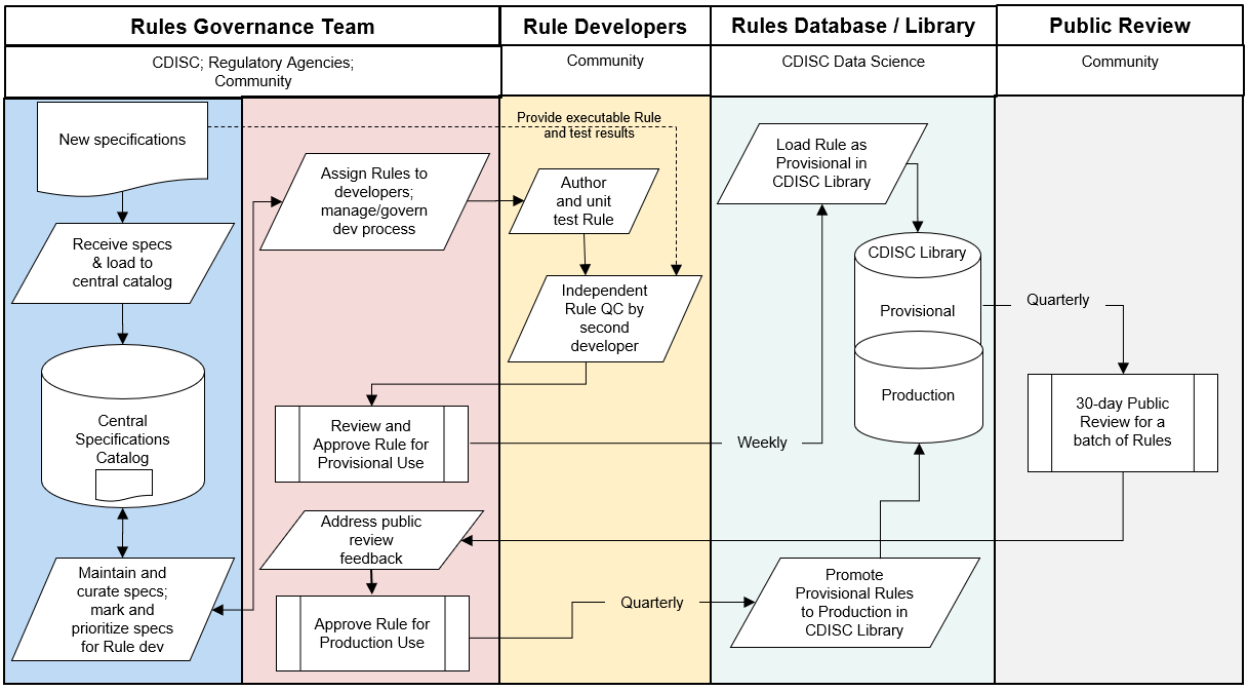
The origin of the rules can be various. As the Foundational Standards come with Rules, all new standards will be containing rules in the future. It is also expected that Regulatory agencies will be creating new rules and the community will come up with rules that are useful for all.

The community (volunteers) will still take on the responsibility of the creation of the rules in the Rule Editor. Once the rules are published, they will become available in the CDISC Library where with the use of an API they can be accessed by the CORE engine and third-party applications.

GOVERNED DEVELOPMENT PROCESS

As shown in Figure 4, the creation and development of new rules will be done according COP001 which includes a 30-day public review period.

FIGURE 4: GOVERNED DEVELOPMENT PROCESS



CORE ENGINE AND DEPLOYMENTS

CORE SOFTWARE

Both CORE Engine and the Rule Editor are open-source software that CDISC has made available for free to the CDISC Community via GitHub and which is listed in COSA. It has been released under MIT open-source license. CDISC will continue to lead the development of both Engine and Rule Editor in the future. Updates can be requested by going to

the cdisc-org repository in GitHub. Extensions of both Engine and Rule Editor are possible as they support the development of software extensions.

The CORE Engine executes CORE Rules (written in YAML) against clinical data and returns results. The code is written in Python and uses the Venmo Business Rule Engine. The Rule Editor is written in TypeScript and uses the VSCode editor.

RUNNING THE CORE ENGINE

The source code of CORE is available on the GitHub repository. The latest release is version 0.6.4 which can found at this link. A CLI (Command Line Interface) is available which allows you to run the rules under Windows, Mac and Linux. The package can be downloaded, unzipped, and run. The GitHub repository will provide all necessary documentation to run a validation, the only thing needed are datasets. The downloaded package already contains a cache with all available rules from the CDISC Library. At any moment, the cache can be updated to get the latest set of rules from the Library.

If you want to run the Engine in your own environment or tooling, it can be implemented as it is available on PyPi. There are also desktop versions available that are created by third-party vendors. More information can be found in the next chapter.

With the original development of CORE (more info can be found in PHUSE EU Connect 2022 DS01 and SD01) a Proof-of-Concept UI was created which shows how Technology providers and third-party providers could run the rules. A short demonstration of this tooling is available at https://www.cdisc.org/core on the "CORE on GitHub" tab.

THIRD-PARTY DESKTOP DEPLOYMENTS

CORE Engine deployments are the domain of the greater CDISC community, including commercial software vendors. Deployments may include those prepared by and for the user community and those prepared by commercial vendors as offerings for the user community. Current deployments:

- Are simple to install and use
- Provide a UI
- Will make it easier for the CDISC community to evaluate CORE without IT support.

CORE deployments must be validated by the deployer. This validation process is separate from CORE Engine validation which is done by the open-source community.

The first free, publicly available vendor-provided CORE desktop version was announced at the CDISC European Interchange. Other software vendors committed to creation a free desktop version or are planning to incorporate it for free in their internal tooling and make it part of the standardized process.

The goal is CORE to be the reference engine which is a commonly used principle in the software industry. It provides a concreate example on how the standard should be implemented. Of course, if software vendors want to use their own engine, that can also be done. In that case, they can get certified by CDISC.

CORE REGISTERED SOLUTION PROVIDER

The purpose of the program is twofold:

- For CORE vendors it is a way to certify that their solutions correctly use the Conformance Rules
- For CDISC it is a way to treat all CORE vendors equally, achieve a level playing field regarding use of any Engine with the Conformance Rules and inform the community which solution has been certified.

The testing for certification will include the generation of results with Conformance Rules and test study data reflecting an "average study". However, CDISC will not check for system functionality as this can differ from each system and can be a distinguisher.

WHAT'S NEXT

On January 16th a press release was published for the collaboration between CDISC and FDA. CDISC and FDA will work together to develop executable formats of the FDA Business Rules and on the development and ongoing governance of this set of executable rules within CORE that can be used by industry (see Figure 5). Having a single version of a rule specification followed by a single version of an executable rule will allow users to see the data in a unified way instead of an interpretation.

FIGURE 5: PRESS RLEASE OF RESEARCH COLLABORATION AGREEMENT

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).

CDISC's CORE project provides an open-source version of the CDISC Conformance Rules in a machine-executable format. These rules, published and managed by CDISC, create a single source for conformance rules and allow external vendors and sponsor companies to implement and extend these rules within their tools. **FDA Business Rules** are currently written in a plain text, non-machine executable format and describe the business requirements for regulatory review to help ensure that clinical trial study data is compliant and useful and supports meaningful review and analysis.

The goal of this effort, which began on November 3, 2023 and has term of three (3) years, is to collaborate on providing input on machine-executable formats of the FDA Business Rules and on the development and ongoing governance of this set of executable rules within CORE that can be used by sponsors of medical product applications.

"CDISC is grateful for the opportunity to partner with CDER and CBER to establish a single source for machine-executable Conformance Rules and drive the implementation by the industry," said Peter Van Reusel, Chief Standards Officer at CDISC.

The benefits of creating a single comprehensive and credible source of validation rules include increasing access, transparency, and visibility of validation rules used to ensure the quality and usability of study data in FDA. This will enable sponsors to submit high quality study data that will be ready for regulatory review saving time and effort for all parties involved.

"Our research collaboration with CDISC is an important step to ensure that study data validation rules are understandable and accessible to all," said Lilliam Rosario, Ph.D., Director, Office of Computational Science, Office of Translational Sciences, CDER.

ADOPTION BY THE INDUSTRY

Besides the Foundational Standards, there are other projects and standards that see additional benefits of the CORE Rule and schema and that moving forward with implementation. A new phase of DDF (Digital Data Flow) has kicked off. In Phase 3a, the CORE team will collaborate with TransCelerate Biopharma to create a Proof-Of-Concept to create and run Conformance Rules on the Unified Study Definition Model (USDM) version 2.6.

On Tuesday 17 of October the Tobacco Implementation Guide (TIG) went out for Public Review. TIG is the first non-Foundational Standard that will also include Conformance Rules. The TIG team has defined a new set of rules but have investigated existing rules and aligned with those.

There are various vendors, sponsors and CROs that have started implementing CORE into their own systems and started the creation of their company-specific (Quality) rules.

NEXT MILESTONE

The next goal for CDISC is to complete the ruleset for SDTMIG 3.3 and SDTMIG 3.4, including Define-XML crosscheck rules. Also high on the priority list are the FDA validator rules V1.6 (that also apply to SDTM IG 3.2 and 3.3) and the FDA Technical Rejection Criteria. The finalization of this ruleset is relying on community engagement to develop these rules. If you are still hesitating to join the CORE development team, read Paper SI05.

The CORE Engine will be released as stable releases so all of the rulesets mentioned above will be able to be executed. The engine always goes through thorough testing and validation documentation is being made available.

The overall purpose is to test the Engine and the Rules with real study data and to roll out the rules governance process.

CONCLUSION

The CORE Program, in the open-source environment today, is a combination of Rules development by the CDISC Standards team, Engine development by the open-source community, and Engine deployment by end-users and by commercial vendors for end-users. A full set of executable rules for submission standards is being created, together with regulatory rules. The CORE engine will serve as the reference engine with all basic functionalities for a full set of machine-executable rules which includes a validation package. When completed, CDISC will establish a CORE Certification Program to verify the output of different applications versus the CORE Reference Engine.

When the Rules and Engine components have been achieved, the CORE Program will provide submission-ready CORE to Industry. At that point, the Rules will be a complete part of the CDISC Standards, and it is expected that Regulatory Agencies will mandate use of the CDISC Conformance Rules.

CDISC Conformance rules are the single version of the truth.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Peter Van Reusel
CDISC
pvanreusel@cdisc.org
www.cdisc.org

Brand and product names are trademarks of their respective companies.