

CDISC Conformance rules and the CORE Engine

Progress and Roadmap

Peter Van Reusel, CDISC, Austin TX, USA

ABSTRACT

This presentation will provide an overview and an update on the CDISC Conformance Rules Engine (CORE). The overview will discuss the CORE concept, roadmap, and landscape of CORE solutions. The update will further discuss industry participation and timelines of the CORE project and the development of the CDISC Conformance Rules.

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.

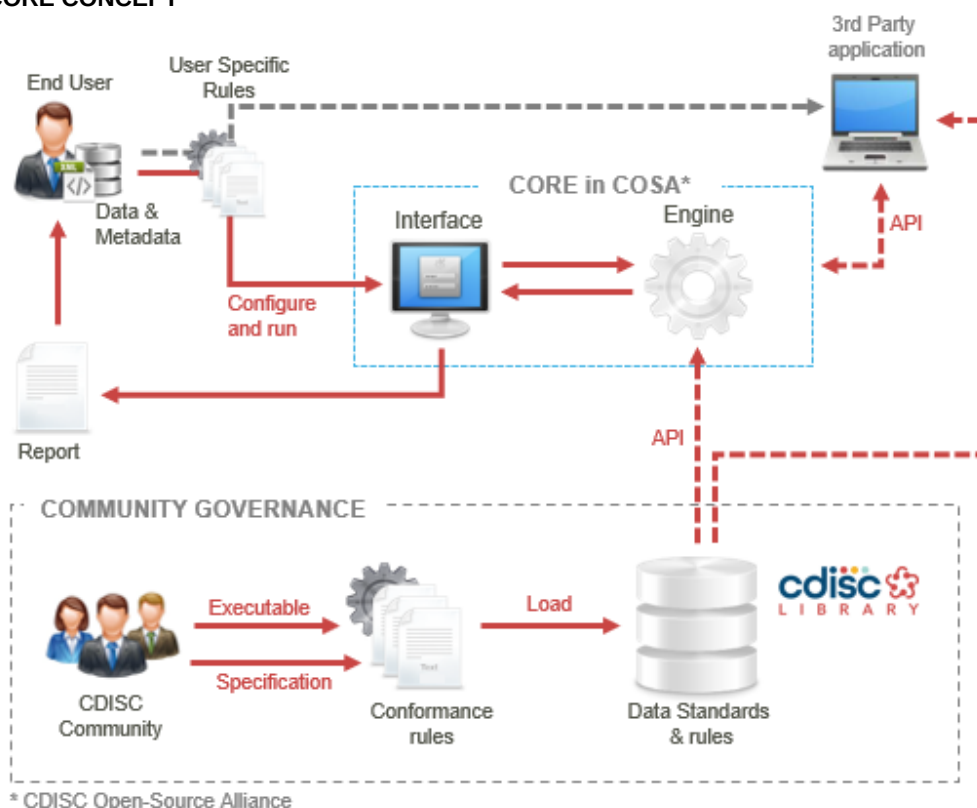
CDISC is conducting the CORE Program for the following reasons:

- 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



HIGH-LEVEL STATUS AND ROADMAP

STATUS

The CORE Program was announced to the CDISC Community at the CDISC 2021 EU Interchange while initial planning was underway. Following is a high-level description of the major activities and milestones to date:

- 2021 JUL –
 - Program kicked off with a Call for Volunteers webinar.
 - Microsoft engaged as a key collaborator on the Engine development.
 - Sprint 0 began of the agile-scrum approach for Engine development.
 - Development of the Rule Editor for creating executable Rules began, including definition of the YAML schema used to express executable Rules.
 - Rules development began with volunteer recruitment followed by actual development; first focus was on the Rules for SDTMIG 3.4.
- 2022 APR –
 - An evaluation version of the Engine was completed and provided to the CDISC Community in the CDISC Cloud Evaluation deployment. The deployment was announced at the CDISC 2022 EU Interchange.
 - 210 of 336 executable Rules for SDTMIG 3.4 were completed and available for the CDISC Cloud Evaluation deployment.
- 2022 JUL –
 - A second deployment of the evaluation version of the Engine was completed and provided to the CDISC Community in the Azure Marketplace. This new cloud CORE deployment showcases how CORE can be deployed and used securely in the cloud as an enterprise solution.

- 2022 Sep –
 - The CORE Program was transitioned to open-source. This included provision of the Engine in GitHub and startup of the CORE Roadmap Board to guide further development of the Engine in an open-source environment. (The CORE Roadmap Board is discussed further later in this paper).
 - Rules development was initiated for additional CDISC Foundational Standards: SDTMIG 3.2 and 3.3; SENDIG; ADaMIG.

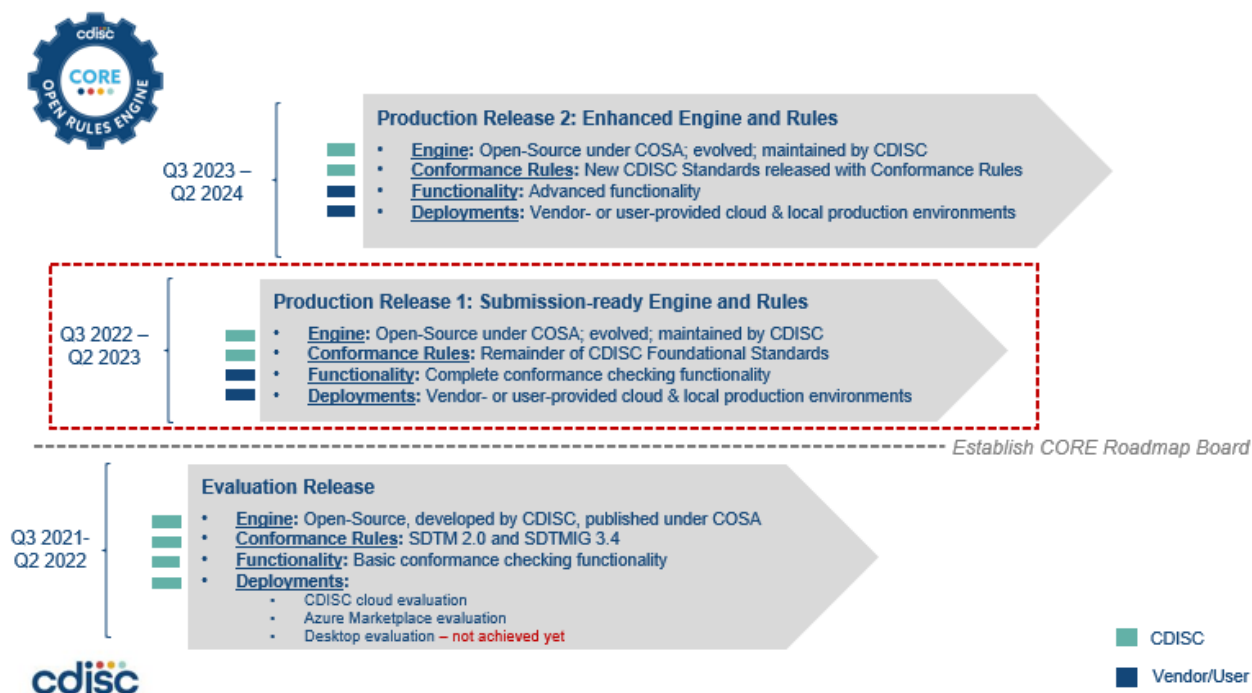
ROADMAP

The CORE Program Roadmap is described in Figure 2 below, with three top-level phases. Phase 1 is complete and focused on development of an evaluation version of the Engine with basic conformance checking capability and on initial Rules development. (The milestones were highlighted immediately above). The engine was deployed in two environments (CDISC Cloud and Azure Marketplace). The CDISC Community began to evaluate CORE, especially in the CDISC Cloud environment. CDISC provided a substantial subset of the SDTMIG 3.4 executable Rules, and a set of test clinical study data, in this environment. A desktop deployment of the evaluation engine was also planned but was pushed back to Phase 2. All activities of Phase 1 were conducted by CDISC as a traditional CDISC project, with volunteer help from the CDISC Community.

Phase 2 is currently underway and is highlighted by transition of the CORE Program from a traditional CDISC project to the open-source environment. The CORE Roadmap Board has been established to guide all activities in this new environment. In the open-source world: (1) the open-source community will be responsible for development of the Engine and the Engine will be a part of the CDISC Open-Source Alliance (COSA), (2) CDISC will be responsible for Rules development with critical volunteer help from the CDISC Community, and (3) Engine deployments will be the domain of end-users and commercial software vendors. The Engine will move from the current evaluation version to a stable version which will be the submission-ready Reference Implementation. The inventory of executable Rules will expand to cover the rest of SDTM as well as many/most of the remaining CDISC Foundational Standards.

In Phase 3, the domains of the respective stakeholders will be the same as in Phase 2. The Reference Implementation of the Engine will mature as will the available deployments. The Rules will be completed for the remaining Foundational Standards and CDISC will be set up to provide new Rules in a timely manner as new Standards are developed.

FIGURE 2: CORE PROGRAM ROADMAP



CORE 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.

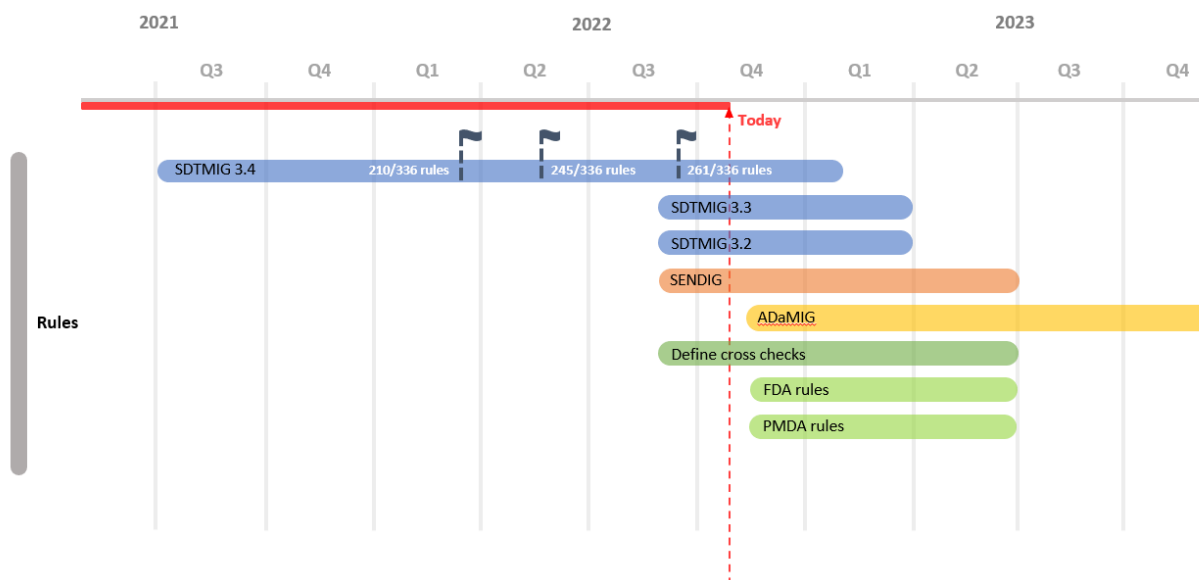
RULE COUNT BY STANDARD

As might be expected, there will be many Rules involved to test conformance to the various CDISC Foundation Standards. Figure 3 provides a summary of the Rule count across the Standards. The Rule set for each Foundation Standards is comprised of two types: those that are executable and those that are not. The breakdown by type is known today for the SDTMIG 3.4 Standard because of the Rule development already done for this Standards. The breakdown by type for the other Standards is yet to be determined. Note in Figure 3 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 DEVELOPMENT TIMELINE

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 prepared in Figure 3 a first-cut timeline for the CORE Program to use as a target.

FIGURE 3: RULES DEVELOPMENT TIMELINE



RULE DEVELOPER SKILL SET, ONBOARDING, AND SUPPORT

A good number of volunteer Rule developers from the CDISC Community will be critical to timely delivery of the Rules. To ensure developer success (and avoid attrition), CDISC has carefully prepared an onboarding and support process for Rule developers. It begins with a description of the skill set for a developer. Interesting to note in this skill

set is that YAML experience is not a needed core skill by only a “plus” skill., A willingness to persevere while working with this new syntax is the ultimate driver of success for a Rule developer:

- Core skills:
 - Data savvy with a science background; e.g., statistics, biometrics, data science
 - A CDISC standards practitioner. Solid implementation experience with SEND, SDTM, and/or ADaM
 - Experience in data specifications & associated verification & validation tasks
- “Plus” skills:
 - Some familiarity with the associated conformance rules
 - Knowledgeable in structured data, such as XML, JSON, YAML
 - A member of an organizational standards council or governance body

The experience of the initial Rule development has shown that success also comes from a solid onboarding and ongoing support for developers. Thus, CDISC has prepared the following carefully crafted onboarding and support program for Rule developers:

- Training webinars
- Training tools on the wiki
- Rule workshops at F2F meetings
- Weekly 1-hour CORE Rule Developer meetings
- Weekly 2-hour CORE Rule Developer “Office Hours”

The CORE Program is currently in need of more Rule developers. Those interested in volunteering can find more information at the “Participate” tab on the CDISC CORE webpage: <https://www.cdisc.org/core>.

CORE ENGINE AND DEPLOYMENTS

ENGINE

The CORE Engine executes CORE Rules (written in YAML) against clinical data and returns results. The evaluation version available today has all the basic conformance-checking functionality. It is open-source software that CDISC has made available for free to the CDISC Community via GitHub.

The Engine is deployment agnostic and may be deployed in the user’s preferred environment (e.g., cloud, on-premises standalone/desktop, on-premises server, etc.). The Engine comes with a command line interface (CLI) for flexible deployment and integration into user systems. CDISC will also provide a basic user interface (UI) to optionally control Engine execution.

Further Engine development has transitioned from a traditional CDISC project to the open-source environment during Q3 2022 and further Engine development is the domain of the open-source community. The open-source environment is characterized as follows:

- Guided by the CORE Roadmap Board and CORE Technical Committee (see below for more)
- Listed in the COSA (CDISC Open-Source Alliance) directory
- Operating under a permissive MIT open-source license

Following is the current focus of Engine development:

- Process new YAML operators added to express new Rules
- Process new clinical data formats
- Support enhanced reporting

ENGINE DEPLOYMENT

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. Core deployments may include:

- An enhanced UI
- Enhanced reporting and issue tracking
- Support for additional clinical data formats
- Ongoing support (e.g., service level agreement)

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.

Currently, the CORE team has made two deployments of the evaluation version of the Engine available to the CDISC Community. A third deployment is planned for the near-term. These three deployments are characterized in Figure 4. To date, the CDISC Community has focused its attention on the CDISC Cloud deployment. CDISC has provided a

subset of the SDTMIG Rules for use with this deployment as well as a set of test clinical study data. Ninety individuals in the CDISC Community have executed 224 validation runs using this deployment.

FIGURE 4: CORE ENGINE DEPLOYMENTS

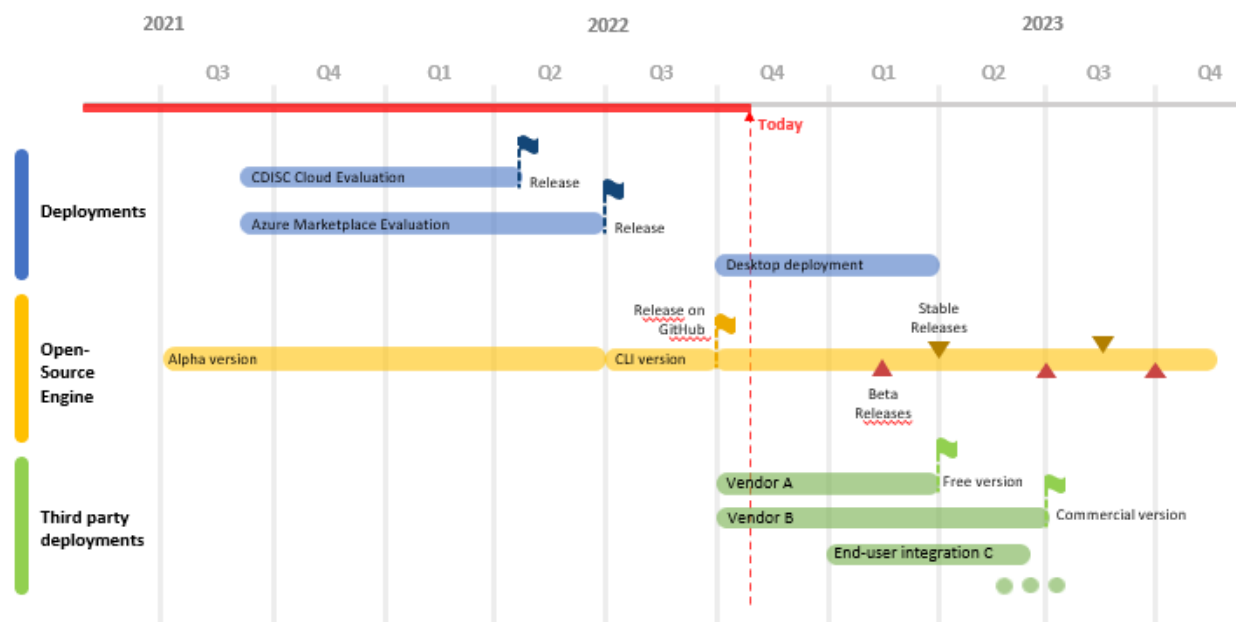
	Available		Planned
	CDISC Cloud Evaluation	Azure Marketplace Evaluation	Desktop
Feature			
Engine evaluation version	✓	✓	Current Engine version
Cloud Deployment	✓	✓	X
Standards Rules from CDISC	✓	✓	✓
User Rules can be added	X	✓	✓
Test data from CDISC	✓	X	X
User data can be added	X	✓	✓
Validate multiple studies concurrently	X	✓	✓
Secure Data Handling	N/A	✓	Depends upon infrastructure
Free to use	✓	X	✓

ENGINE AND DEPLOYMENT TIMELINE

The timeline for Engine and Deployment development is described in Figure 5. As stated above, previous Engine development in the CORE Program has focused on the evaluation version, which has been deployed in two cloud-based environments. A third deployment (desktop) is planned for delivery during Q1 2023. The desktop deployment will involve the Engine evaluation version, or the first stable Engine Reference Implementation if it is available at that time.

Currently, CDISC is aware of two in-development third-party commercial deployments of the Engine. Their timeline is shown in Figure 5 as described to CDISC by the respective vendors. Another third-party bar is shown in Figure 5 to represent the end-user deployments expected by CDISC per discussions to date with the CDISC Community. The dates for these are unknown, as represented by the 3 dots near this bar.

FIGURE 5: ENGINE AND DEPLOYMENTS TIMELINE



CORE ROADMAP BOARD

As stated above, during Q3 2022 the CORE Program has transitioned from a traditional CDISC project to an open-source environment. In this transition: (1) CDISC has made the Engine software available to the CDISC Community for free in GitHub, (2) Engine development is now the responsibility of the open-source community, (3) a CORE Roadmap Board has been established to guide the CORE Program, and (4) Rule development by CDISC will be in coordination with the CORE Roadmap Board.

The overall landscape for CORE development and participation in the open-source environment is illustrated in Figure 6. Central to that landscape is the CORE Roadmap Board. Per the CORE Roadmap Board Charter, membership on the Board is open to those most-engaged Industry stakeholders from Pharma-Biotech, Independent Consultants, and Service Providers. CDISC will also be represented on the Board. An avenue for commercial software vendors to interact with the Board is provided through the CORE Registered Solution providers program to be developed.

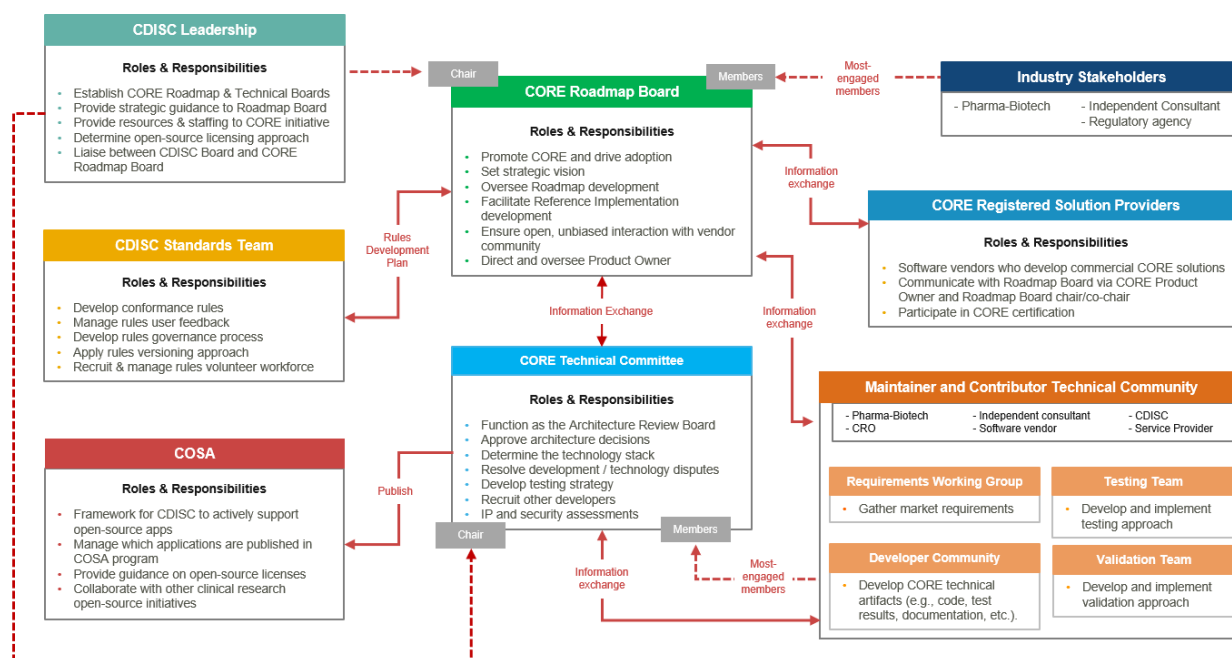
Highlights of the CORE Roadmap Board responsibilities are to:

- Promote CORE and drive its adoption
- Set the CORE strategic vision
- Oversee CORE Roadmap development
- Facilitate the Engine Reference Implementation development
- Ensure open, unbiased interaction with the vendor community

Other CORE stakeholders in this landscape are:

- CDISC Leadership, which will be represented on the Board and provide Board leadership in year 1
- CDISC Standards Team, which will further develop the Rules in coordination with the Board
- CDISC Open-Source Alliance (COSA), of which CORE will be a listed application
- CORE Technical Committee, which will be comprised of the most-engaged members of the open-source community who are involved in the Engine development
- Maintainers and Contributors in the open-source community, who are the other stakeholders who are actively involved in Engine development
- CORE Registered Solution Providers, who develop commercial CORE solutions

FIGURE 5: LANDSCAPE OF CORE DEVELOPMENT AND PARTICIPATION



NEXT STEPS

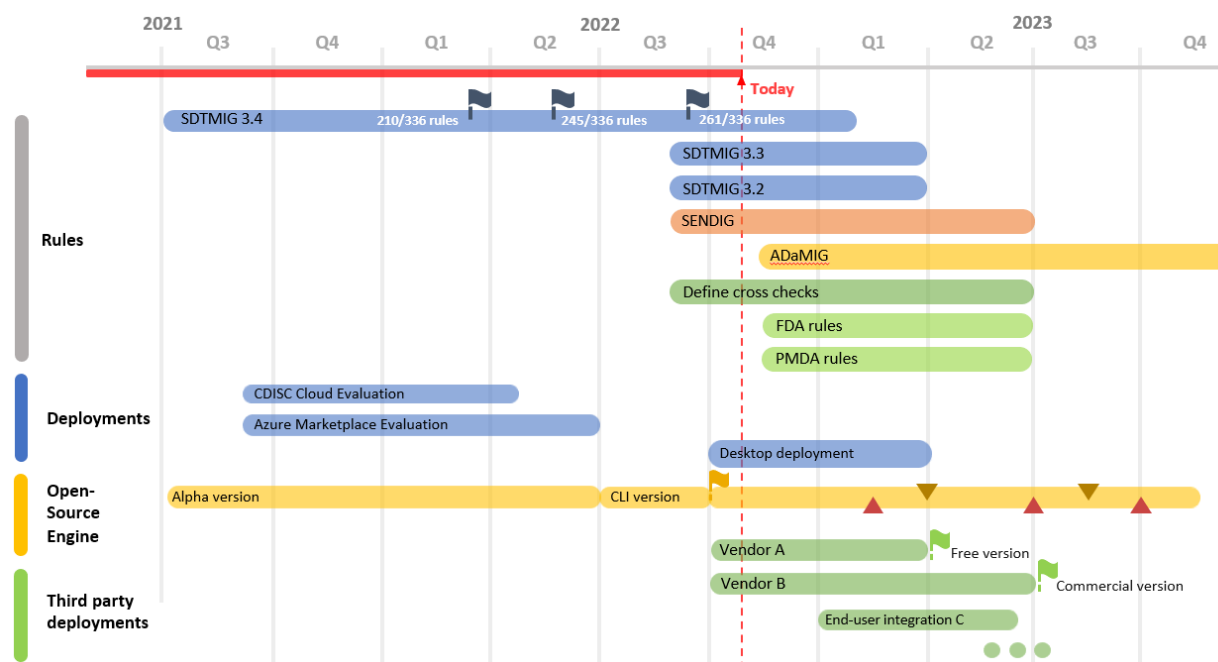
At the highest level, the next step for the CORE Program is to provide “submission-ready CORE” to Industry, where submission-ready CORE is described as follows:

- A full set of Rules
 - A validated set of rules for submission standards (SDTM, SEND, ADaM)
 - Including Regulatory-specific rules
 - Including Define.xml cross-check rules
- CORE is the Reference Engine
 - An Engine with all basic functionality for the full set of machine-executable rules
 - Includes a validation package
- A CORE certification program
 - To verify output of different conformance applications versus the CORE Reference Engine
 - CDISC conformance Rules are the single version of the truth

When the CORE Program achieves submission-ready CORE, the Rules will be a complete part of the CDISC Standards. It is expected at that time that Regulatory Agencies will mandate use of the CDISC Conformance Rules.

The consolidated time for the CORE Program, across the Rules, Engine, and Deployment efforts is described in Figure 6. As was stated above, there are challenges to establishing a Program timeline and certain assumptions had to be made, including regarding the amount of labor available to the Rules and Engine efforts. Thus, this timeline is considered the best understanding of the CORE Program team today.

FIGURE 6: CONSOLIDATED CORE TIMELINE



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. This paper has described each component and its status and roadmap. The timeline, as it is understood today, for completing these three components was presented in Figure 6.

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.

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.