

Paper SI03

Our Road to Adopting CORE: an Implementation Journey in SDTM Dataset Validation

Marisa Wyckmans, SGS Health Science, Mechelen, Belgium
Els Janssens, SGS Health Science, Mechelen, Belgium

ABSTRACT

The Secure Data team, an unblinded team within SGS Data Management, daily handles numerous data transfers. Particularly during critical periods, delivery of SDTM compliant datasets is defined by a tight turnaround time. Consequently, the team relies on tools that can convert and validate high quality datasets efficiently. We are enhancing our dataset validation tool and combine this with our experience and knowledge from the CDISC Open Rules project. This presentation illustrates the requirements for our new dataset validation tool, including the transition to the cloud, and aligning them with the capabilities of CDISC Open Rules Engine (CORE). Moreover, we showcase that we can create our own rules using the CORE Rule Developer and discuss the challenges we faced. We are also outlining and finetuning our own SGS pathway for future requirements/implementations. In collaboration with CDISC, we're making significant strides towards establishing an open, single source of truth for conformance rules and consistency.

INTRODUCTION

In our industry, our primary goal is to generate clinical datasets that are submission ready. To achieve this, it is considered best practice to continuously run conformance rules, which are defined by CDISC and regulatory agencies, on the data throughout the study's duration. However, a significant challenge arises from the fact that multiple organizations have their own interpretations of rule descriptions. This divergence in understanding can ultimately result in discrepancies in the data during submissions. Furthermore, a drawback is that the conformance rules, defined by CDISC and regulatory agencies, are not currently executable.

To ensure consistency and accuracy, it is crucial for all stakeholders to speak the same language and rely on a single source of truth. This is where the CDISC Open Rules project comes into play. The overall goal of the CDISC Open Rules 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. By establishing standardized rules and promoting their adoption across the industry, this initiative is very valuable in harmonizing rule descriptions and fostering collaboration among stakeholders. It plays a vital role in facilitating the creation of submission-ready clinical datasets that meet regulatory requirements and align with industry best practices.

CDISC OPEN RULES

Figure 1 below illustrates the CDISC Open Rules concept^{1,2}. The CDISC Open Rules project 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 it is to execute the Rules against clinical data and return results. When the Engine operates, it retrieves the Rules from the CDISC Library via an API. Users may also add custom Rules for processing. 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. Finally, an alternate engine may be developed (by an end-user or by a commercial software vendor) to execute the Rules.

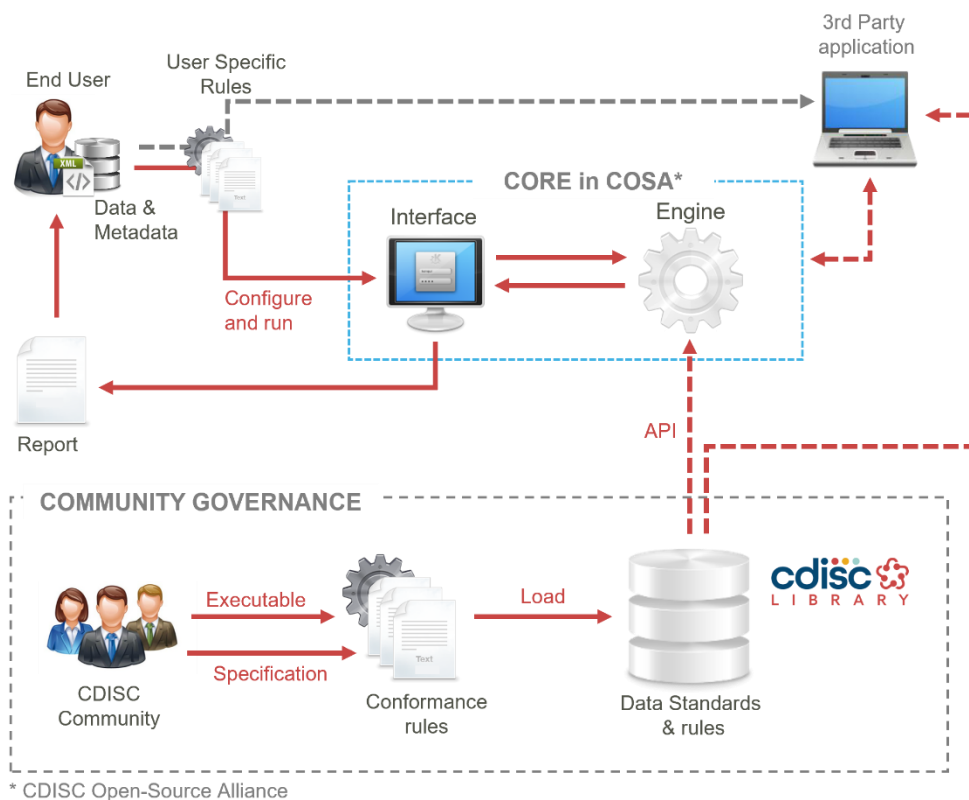


Figure 1: CDISC Open Rules Concept

At SGS, we are highly interested in enhancing our data processing capabilities by incorporating custom rules. In addition to the existing CDISC governed rules, we would like to create and execute our own rules using the CDISC Open Rules Engine. This would enable us to establish our own local rule library and effectively execute these rules against our data. Our custom rules would encompass various aspects such as:

- Rules that are currently not included in the CDISC governed set.
- Data cleaning rules
- Non-CDISC clinical data, including vendor data.

The CDISC Open Rules Engine, an open-source tool for validating datasets, can be seamlessly integrated into our existing infrastructure or utilized as a standalone application. To develop our local rules, we intend to leverage the open-source rule editor provided by CDISC.

The concept of developing an industry standard method for rule creation holds immense potential. It could pave the way for new forms of interaction among various stakeholders. For instance, if a custom rule proves beneficial to the entire community, it could be considered for inclusion in the CDISC-governed set.

We anticipate the possibility of transferring custom rules between different entities, such as pharmaceutical companies (Pharma) and contract research organizations (CROs), or between CROs and vendors. The ability to exchange custom rules between different entities facilitates seamless collaboration and ensures adherence to specific requirements.

In summary, the ability to create local rules opens opportunities for collaboration, standardization, and improved data validation across different stakeholders within the industry.

IMPLEMENTATION JOURNEY

With great enthusiasm, we have been exploring opportunities to implement the concept of custom rules. We have identified a specific use case where this concept can be put into practice effectively - the data validation conducted by our Secure Data (SD) team. Our SD team is our unblinded Data Management (DM) group, with the crucial responsibility of handling unblinding datasets. As part of their role, they perform data validation to ensure accuracy and integrity of the unblinding datasets.

UNBLINDING DATA FLOW

When lab sampling is performed in a clinical trial, it is possible that certain sampling results, such as PK (pharmacokinetics), biomarkers, or immunogenicity, need to be blinded. In such cases, the lab sends these results to

our SD team in various formats. The SD team is responsible for converting these lab transfers into SDTM datasets, such as PC, LB, or IS. These SDTM datasets can then be sent to third parties, such as Data Management (DM) or Statistics, either in a blinded or unblinded manner. This flow is illustrated in Figure 2 below.

To ensure the integrity and compliance of both incoming and outgoing transfers, Data Transfer Agreements (DTAs) are established. These DTAs define the structure and content of the transfers. For incoming transfers, validation is necessary to ensure compliance with the DTA and stability as a reliable source for SDTM conversion. The outgoing transfers on the other hand require validation to ensure compliance with the DTA, adherence to SDTM standards, and alignment with regulatory requirements.

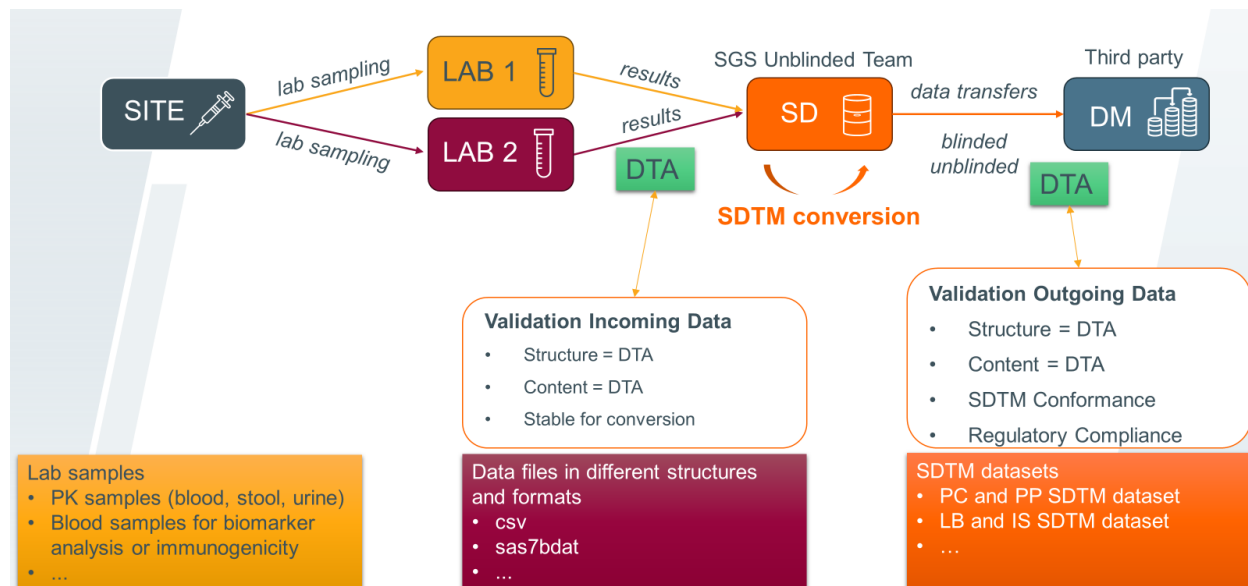


Figure 2: Unblinding Data Flow

A NEW DATA VALIDATION TOOL

We were already in the process of enhancing the current SD Data Validation Tool for following reasons:

- The complex rule output often consists of just a list of data, which requires a lot of manual effort to check data conformance.
- Rule creation requires programming knowledge and extensive documentation.
- New SDTM or regulatory conformance rules need to be added manually, without any automation.
- Rules combine different aspects for validation, making it challenging to manage.
- The tool only validates the outgoing datasets converted to SDTM, while incoming transfers are validated manually.

Our vision for the new Data Validation Tool is:

- Rules should generate simple output. If the data is conformant, no output is expected, reducing unnecessary work.
- Rule creation should be easy, using low code and an efficient workflow.
- Automatic upload of new SDTM or regulatory conformance rules should be supported.
- Rules should be categorized based on their validation purposes for increased comprehensiveness.
- There should be an option to validate both incoming vendor files and outgoing SDTM converted datasets.

During discussions with our small group of three early CDISC Open Rules enthusiasts, we realized that our vision aligns closely with the capabilities of CDISC Open Rules and the custom rule concept. Therefore, we decided to explore the possibility of implementing the CDISC Open Rules vision and related tools in house to enhance our data validation strategy.

To create our custom rules, we needed to deploy the rule editor locally and ensure that the engine could pick up and execute these rules. Our team of programmers dedicated their efforts to this project at the beginning of this year, and by March 2024, had successfully deployed the rule editor in our SGS cloud environment. This allows us to create local rules and execute them with the engine. The next task is to develop the actual rules themselves.

THE SD RULE PACKAGE: CDISC GOVERNED RULES MEET SGS DEFINED RULES

We started creating the new SD rule package from the ground up, based on the checks available in the old validation tool. For each check, we examined the rule's objective and cross-referenced it with the existing CDISC Conformance rules and FDA Business rules to avoid duplicating efforts. Consequently, we then focused on filling in the gaps with

custom rules. This approach resulted in multiple rules stemming from a single check in the previous tool. In the end, we compiled 356 rules in the SD rule package (see Figure 3 below), categorized into SDTM conformance (regulated by CDISC, FDA, or SGS rules), DTA Compliance, and Data Review. Most of these rules fall under SDTM Conformance, and the CDISC Library already provides a convenient source for 73% of the SDTM conformance rules required by the SD team. This is certainly positive news.

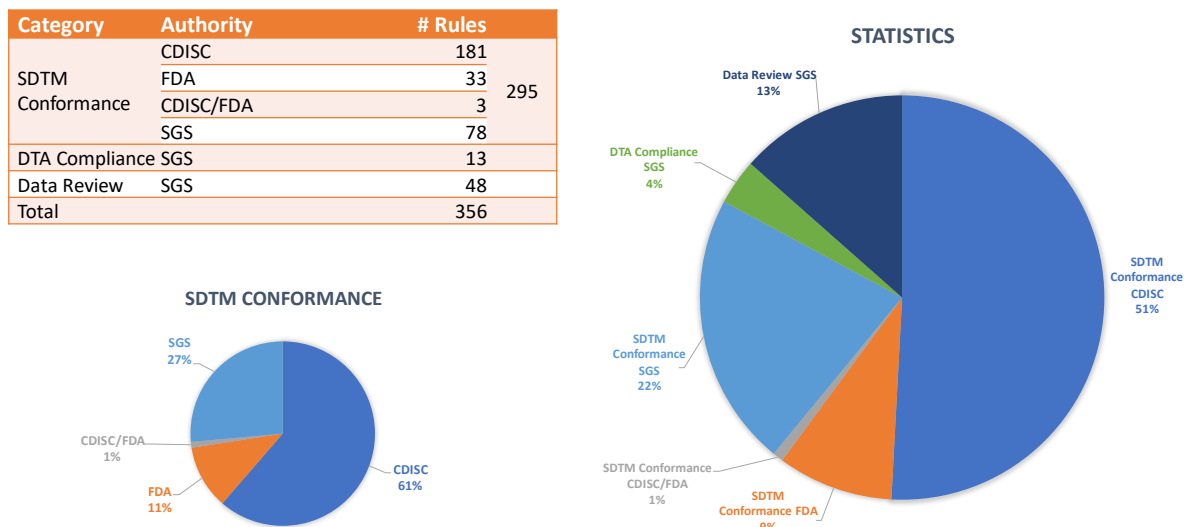


Figure 3: SD Rule Package – categorization into SDTM Conformance, DTA Compliance and Data Review rules.

CUSTOM RULE WRITING EXAMPLES

One specific example from the previous tool involves creating an overview of results and limits of quantitation to ensure alignment. This check is broken down into five custom rules (see Figure 4 below). The first two rules (SGS-000510 and SGS-000511) are simple comparisons, verifying that --STRESN falls within the range of --LLOQ and --ULOQ. The remaining three rules (SGS-000512, SGS-000513 and SGS-000521), however, need to extract a subset of information from the dataset, which then requires manual review to assess if the content meets the necessary criteria.

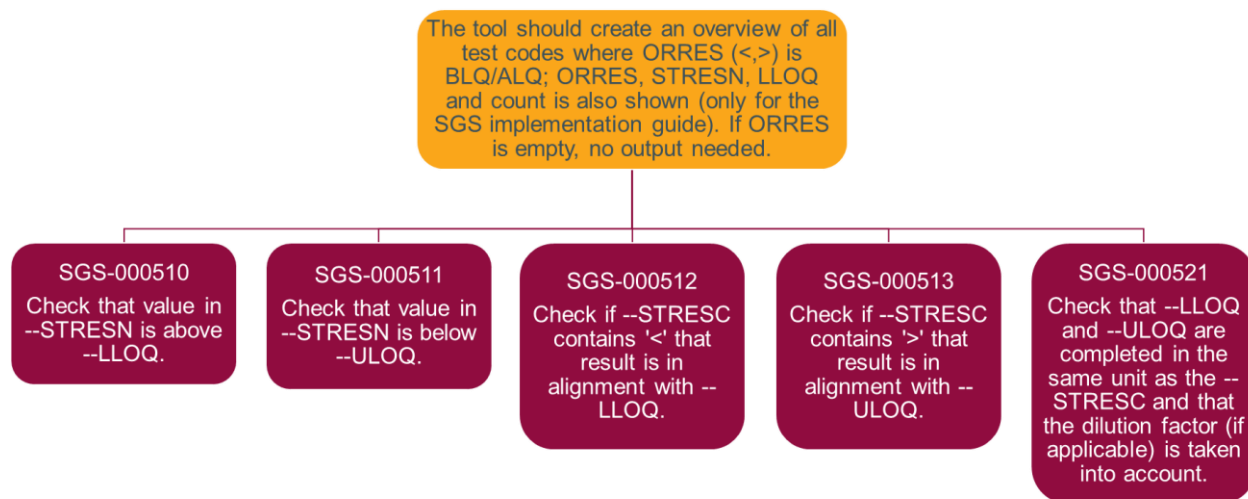


Figure 4: Breakdown of a Data Content check into several SGS Custom rules.

Rule SGS-000512 is not a simple check but rather a listing. The objective is to identify records where --STRESC contains a smaller than sign ('<') and also capture the associated --LLOQ information. This rule can be conveniently written in the rule editor (see Figure 5 below), using rule logic to filter the desired records. We provide a more detailed description and message for this rule because the output will require manual assessment. An important advantage is that we can select output variables based on relevance to the check's purpose or for easy identification of records if further action is needed, even if they are not utilized in the rule logic. This rule is tested with sample data (see Figure 5 below), and it successfully produces the expected records from the dataset (see Figure 5 below). Hence, this demonstrates that we can create rules that go beyond straightforward checks and generate comprehensive listing-like outputs.

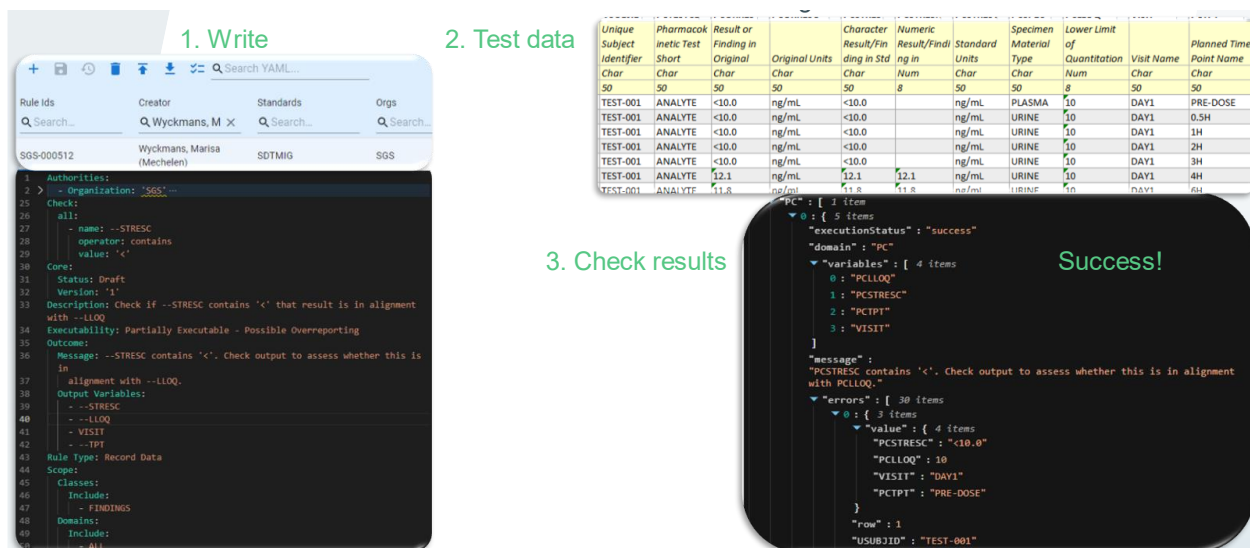


Figure 5: Process of creating a listing for SGS-000512: 1) write the rule logic in the rule editor, 2) create test data, 3) run the rule and check results.

The DTA Compliance rules pose a significant challenge due to the varying structures of DTAs. Amongst others, ensuring adherence to the DTA requires checks to verify that the test information aligns with the DTA specifications (SGS-DTA001), and that the vendor's name is in accordance with the DTA (SGS-DTA002) (see Figure 6 below).

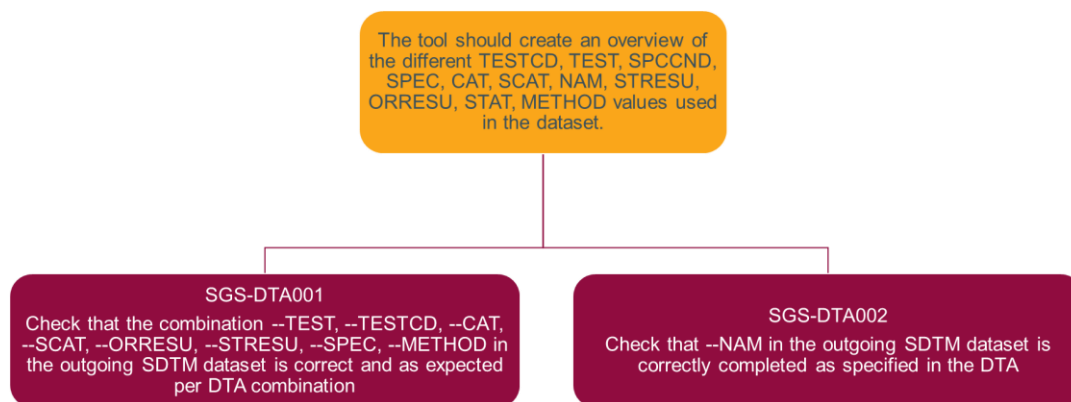


Figure 6: Example of DTA Compliance rules.

As illustrated in Figure 7 below, the rule editor allows to create a rule (SGS-DTA002) to verify the vendor's name against the DTA specifications. Since the DTA specifications table is organized vertically, the value corresponding to the PCNAM field needs to be extracted for this rule. In the rule logic, we use an operation to find the value while filtering based on the PCNAM, as that is the specific field of interest. By comparing this extracted value with the vendor's name in the dataset, we can identify discrepancies. There are two incorrect PCNAMs in the test data for PC (yellow highlights) and this rule successfully flags these inconsistencies and ensures compliance with the DTA requirements.

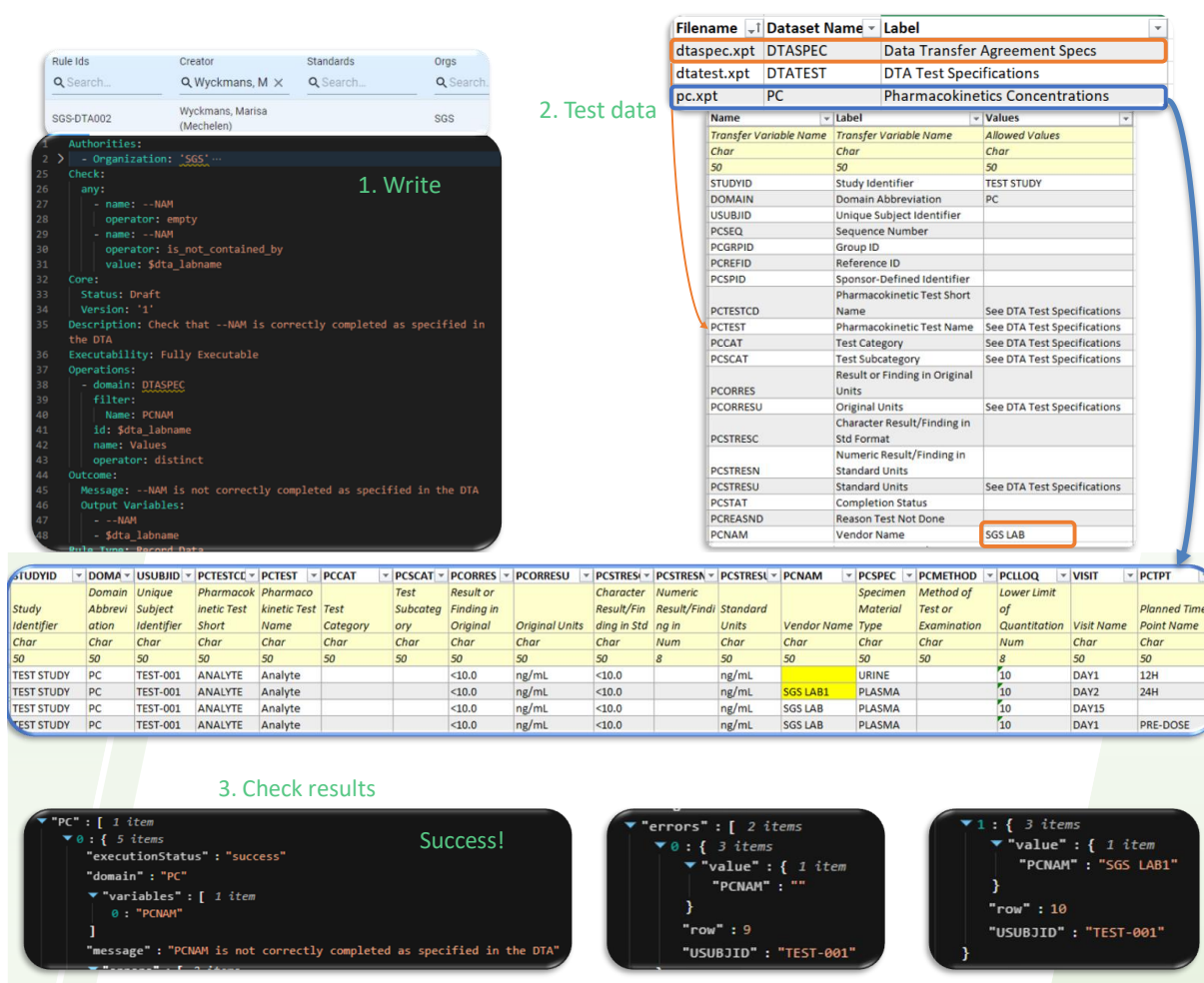


Figure 7: Process of creating DTA Compliance rule SGS-DTA002: 1) write the rule logic in the rule editor, 2) create test data, 3) run the rule and check results.

INSIGHTS SO FAR

Here's what we have learned so far:

- There is no difficulty creating additional SDTM conformance rules using the local rule editor. However, when publishing a rule in the local rule editor, an issue arises with the automatic assignment of Core IDs, resulting in overlaps with CDISC governed rules. To ensure the uniqueness of Core IDs, further investigation is needed to avoid this overlap between CDISC rules and company rules.
- Contrary to initial expectations, writing DTA compliance rules does not require conversion due to the vertical structure of the DTA. This is good news, as it simplifies the process.
- Checking data content is also feasible. However, updates to the schema will be necessary to accommodate company-specific data and parameters. Additionally, new operators may need to be introduced to enhance the capabilities of data content checks.
- It is worth noting that data structures other than the standardized and clean SDTM format are also possible, which expands the possibilities for rule creation and analysis.

Overall, these findings indicate that there are solutions and opportunities for creating and implementing rules beyond SDTM conformance, including DTA compliance and data content checks.

GLIMPSE TO THE FUTURE

Our journey began approximately a year ago, and since then we have successfully programmed, and quality checked 51% of the custom SDTM conformance rules. Although we have not yet started programming data content and DTA compliance rules, our tests indicate that we will be able to accomplish this task. However, there are still challenges ahead of us, particularly with regards to naming conventions and governance. Additionally, we need to ensure that the tools pass CSV validation and establish standard operating procedures (SOPs) and guidelines. While there is still work remaining, we are pleased with our progress thus far and anticipate completing the project next year (see Figure 8 below).

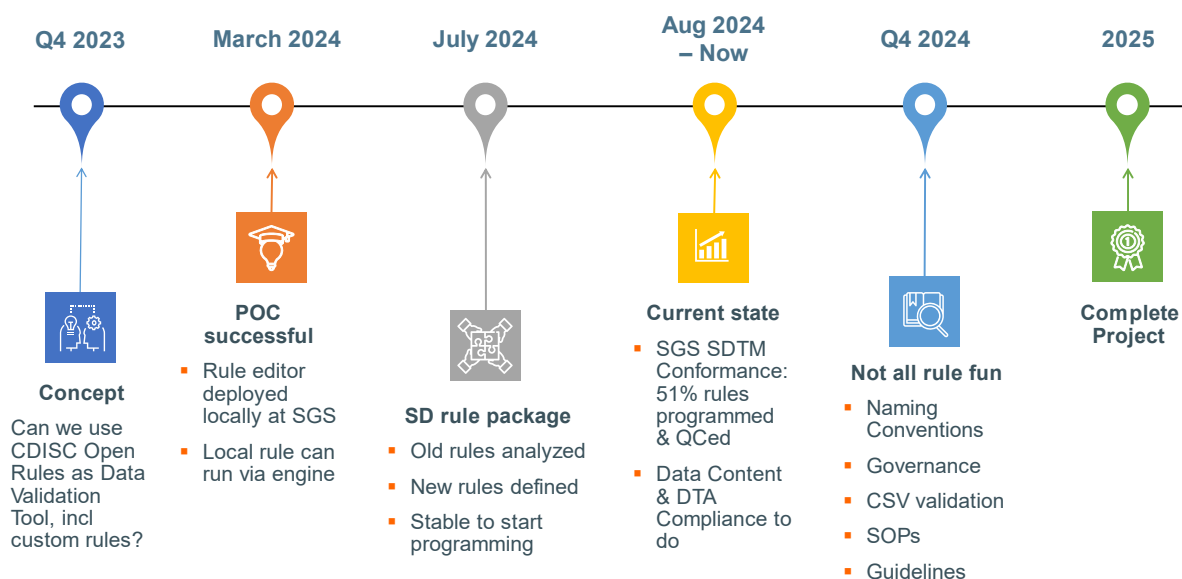


Figure 8: Project Timeline

In addition, there are two important points to highlight. Firstly, to initiate the validation process using the rule engine, certain parameters need to be configured through the command-line interface (CLI). Secondly, the output generated by the engine is in Excel format.

As a non-tech-savvy end user, one may find it challenging to work with the CLI. Moreover, manually comparing different Excel files every time the validation is run is both time-consuming and error prone. Therefore, the vision at SGS is to create more user-friendly platforms, 1) to simplify the configuration of the validation process, and 2) to provide efficient ways to review the output, track changes, and add comments. Both objectives are currently being worked on as separate, ongoing projects.

CONCLUSION

The CDISC Open Rules project aligns perfectly with the dataset validation needs of our SD team, and at SGS, we see great potential in it. It's important to note that our journey is still ongoing, with implementation currently underway. As this implementation project progresses, we may find the need to fine-tune the rule package to ensure completeness and optimal performance. Additionally, we have plans to expand beyond unblinding data and incorporate checks for all types of data.

Furthermore, we are actively engaging with CDISC to discuss further developments of the code, both for the rule editor and the engine. This engagement demonstrates our commitment to continuous improvement and collaboration and will undoubtedly contribute to the advancement of the project.

REFERENCES

- <https://www.cdisc.org/core>
- Nick De Donder & Peter Van Reusel (PHUSE EU Connect 2023 - SI04): CDISC Conformance rules and the CDISC Open Rules Engine Continuing the Road to Adoption

ACKNOWLEDGMENTS

Great things in business are never done by one person. They are done by a team of people. We would like to thank our internal team as well as the CDISC team for their invaluable contributions.

CONTACT INFORMATION

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

Marisa Wyckmans
SGS Health Science

Generaal De Wittelaan 19A bus 5
BE - 2800 Mechelen

Work Phone:

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

Email: clinicalresearch@sgs.com

Web: www.sgs.com/cro

Brand and product names are trademarks of their respective companies.