

rock and roll – Deliverables driven by standardized data specifications

Svetlana Freiberga, Pharmetheus, Uppsala, Sweden
 Mounika Suddapalli, Pharmetheus, Uppsala, Sweden
 Georges Etté, Pharmetheus, Uppsala, Sweden
 Arnaud Gaby, Pharmetheus, Uppsala, Sweden
 Joakim Hellner, Pharmetheus, Uppsala, Sweden
 Pēteris Prūsis, Pharmetheus, Uppsala, Sweden
 Eric Sandström, Pharmetheus, Uppsala, Sweden

ABSTRACT

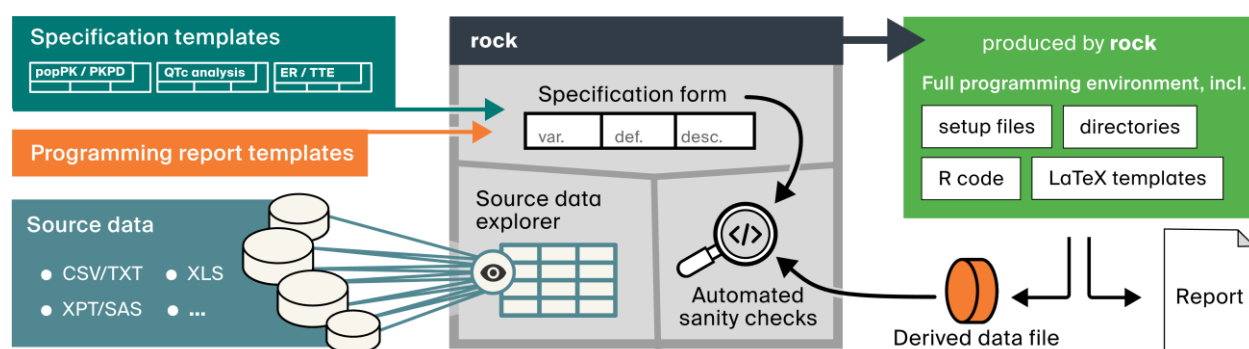
Creating data deliverables from scratch is repetitive, error-prone, and difficult to scale. To support data programmers in this process, while enforcing adherence to standards and maintaining high quality, Pharmetheus has developed **rock**, an R Shiny application. **rock** streamlines everything from drafting data specifications to generating ready-to-use programming environments. It helps standardize workflows and ensures reproducibility and full traceability while retaining flexibility. **rock** brings programmers and modelers together in a shared, version-controlled framework that adapts to evolving project requirements. It's not just about automation; it's about building smart tools people actually want to use. **rock** is a practical solution designed around day-to-day challenges, making workflows more efficient, transparent, and scalable.

INTRODUCTION

The process of creating analysis-ready data deliverables is a vital part of any pharmacometrics workflow, but is prone to being slow, repetitive, and full of manual steps. For every new study, programmers might often be required to rebuild the same fundamentals - specifications, folder structures, and report templates from scratch. This is time-consuming and makes it difficult to adhere to internal or industry standards.

The purpose of this paper is to demonstrate how **rock** helps ensure that project adheres to necessary standards and maintains a high level of quality while allowing for complete reproducibility and traceability. **rock** brings programmers and modelers together within a shared, version-controlled framework. This paper provides an overview of how **rock** achieves its goal by offering a practical solution built around everyday operational challenges.

THE ROCK FRAMEWORK



Specification form - Users can create and customize data specifications using pre-defined templates for various analysis types. Variables can be added, edited, or removed, with detailed metadata including definitions, units, rounding, dependencies, and derivations.

Source data integration - Users can upload and manage different types of source data. The system tracks each variable's origin, including study, file, and variable mappings, ensuring full transparency.

Data programming environment - **rock** generates a structured R/LaTeX environment with predefined directories, R script placeholders, and LaTeX templates for report generation.

Automated sanity checks - Provides quick overview of data and specification consistency.

Version control and traceability - All edits, derivations, and updates are tracked, maintaining a clear audit trail.

DATA SPECIFICATION GENERATION

The process of data specification generation within **rock** is built around standardized templates tailored to different types of analyses such as PopPK, PKPD, QTc, TTE. These templates serve as the essential starting point for defining data requirements, ensuring a consistent and efficient approach. By leveraging standardized specifications, **rock** streamlines workflow and establishes a foundation based on predefined rules, reducing the risk of errors that arise when creating specifications from scratch.

The interface shows the 'Source' tab with a navigation bar at the top. Below the navigation bar, the version information 'PMXRock version: 2.0.12 encore version: 2.0.11' is displayed. The main area is titled 'Choose the project and master/template file, then click Load file'. It contains two dropdown menus: 'Client folder' (set to 'Pharmetheus') and 'Client project' (set to 'PMX-CONF-PMX-1'). To the right, a 'DDT master file' dropdown is open, showing a list of files including 'rocksource-1a-PMX-CONF-PMX-1.csv' and several 'Default templates'. A 'Load file' button is to the right of the dropdown. Below the dropdowns, there is a 'Create DDT preview PDF' button. At the bottom, there is a 'Variable' and 'Definition' section with a 'Show 15 entries' dropdown and 'Previous', '1', 'Next' navigation buttons. A 'Quit' button is in the bottom right corner.

DATA SPECIFICATION ADJUSTMENT

The interface shows the 'Source' tab with a navigation bar at the top. Below the navigation bar, the version information 'PMXRock version: 2.0.12 encore version: 2.0.11' is displayed. The main area is titled 'Pharmetheus/PMX-CONF-PMX-1' and 'rocksource-1a-PMX-CONF-PMX-1.csv'. It contains a 'Create DDT preview PDF' button, a 'Save As' button, and a 'Save' button. Below these, there is a 'Search:' input field. The main table has three columns: 'Variable', 'Definition', and 'Comment'. The table contains 15 rows of variables, including 'REF', 'STUDYIDN', 'ID', 'TYPE', 'EVID', 'DV', 'BLQ', 'LNDV', 'LLOQ', 'SEXN', 'RACEN', 'AGEYBL', 'USUBJID', 'DATETIME', 'AMT', and 'NONMEM dose amount administered'. To the right of the table, there is a sidebar with buttons: 'New empty variable', 'Change variable order', 'Include/Exclude multiple', 'Show included only', and 'Binarize variable'. At the bottom, there is a 'Show 15 entries' dropdown and 'Previous', '1', '2', '3', '4', '5', '...', '8', 'Next' navigation buttons. A 'Quit' button is in the bottom right corner.

Variable	Definition	Comment
REF	Unique record number	
STUDYIDN	Study identifier	
ID	Unique subject identifier within current file	
TYPE	Type of record	
EVID	NONMEM event identification	
DV	NONMEM dependent variable	
BLQ	Flag for concentration reported as below $\text{gl}_{\text{S}}(\text{LLOQ})$ in source data	
LNDV	Log-transformed dependent variable	
LLOQ	$\text{gl}_{\text{S}}(\text{LLOQ})$	
SEXN	Sex	
RACEN	Race	
AGEYBL	Age baseline value in years	
USUBJID	Unique subject identifier across studies as provided by source	
DATETIME	Date and time from source file	
AMT	NONMEM dose amount administered	

The data specification editing phase provides the flexibility needed to tailor standardized specifications to project-specific requirements. During this stage, users have full control over the variables included in the specification. Standardized variables that are not relevant to the analysis can be unselected, while additional standardized variables can be added from the template.

To accommodate unique analysis needs that fall outside standard templates, new non-standard variables can also be manually defined and added. This ability to combine standardized and custom variables ensures that the framework remains highly flexible while maintaining consistency and adherence to standards. Ultimately, this enables the specification to evolve alongside project requirements, ensuring the documentation accurately reflects the data necessary for modeling and analysis.

Variables Source Source viewer DAT vs DDT LaTeX macros Global info Export Help Preferences

PMXRock version: 2.0.12 encore version: 2.0.11

Create DDT preview PDF

Pharmetheus/PMX-CONF-PMX-1
rocksource-1a-PMX-CONF-PMX-1.csv
DDT-1a-PMX-CONF-PMX-1.csv

Save As Save

Search:

Variable

- REF
- STUDYIDN
- ID
- TYPE
- EVID
- DV**
- BLQ
- LNDV
- LLOQ
- AMT

Cancel Delete variable Edit existing variable Save as new variable OK

☒ Include in DDT

Variable DDT-id 7 DV

Definition NONMEM dependent variable

Unit ug/ml

Dependencies on other variables?

Possible Values

Description?

Rounding To Decimals

Not present in data

EVID==0 & BLQ==0	Numeric	Observed value.	No rounding	☐
EVID==0 & BLQ==1	LLOQ/2	Set to the 'glc(LLOQ)' divided by 2.	No rounding	☐
EVID==1	0		No rounding	☐
Can be empty			No rounding	☐

The editing features available during the data specification adjustment allow users to precisely define or modify key variable attributes. Upon selecting a variable, users can edit its name and definition, as well as specify or update attributes such as possible values and descriptive context, if needed. For quantitative variables, the required unit of measure and rounding rules can be clearly defined. An essential technical feature at this stage is the ability to define programming dependencies, ensuring that all computational prerequisites are properly documented.

Derivation rules

Derivation (Optional Internal)

Requirement:

For ALQ=1 need to communicate with the project team if DV can be set to ULOQ or other approaches need to be used.

In the source column always need to specify which PK type of records are selected from source, e.g.,

Derived from pc.sas7bdat: PCSTRESN, PCLLOQ

where PCTEST="C0125" and PCSPEC="PLASMA".

In addition, this tab provides immediate guidance to programmers responsible for implementation and facilitates collaborative documentation, including notes or event comments from the project team members.

THE SOURCE VIEWER

[Variables](#)
[Source](#)
[Source viewer](#)
[DAT vs DDT](#)
[LaTeX macros](#)
[Global info](#)
[Export](#)
[Help](#)
[Preferences](#)

STUDYID : CDISCPLOT01
DOMAIN : PC
PCTESTCD: XAN
PCTEST : XANOMELINE
PCORRESU: ug/ml
PCSTRESU: ug/ml
PCNAM : Imaginary Labs
PCSPEC : PLASMA
PCLLOQ : 0.01
VISIT : BASELINE
VISITNUM: 3

Favorite source files

20250827/ST1/pc.sas7bdat

View as table

Columns to display (comma-separated, empty field means show all columns)

USUBJID,PCSEQ,PCORRES,PCSTRESC,PCSTRESN,PCDTC,PCDY,PCTPT,PCTPTNUM

show N records

500

Update

Compute summary information for filtered data

Hide summary

Show

10

 entries

Search:

row	USUBJID	PCSEQ	PCORRES	PCSTRESC	PCSTRESN	PCDTC	PCDY	PCTPT	PCTPTNUM
row	[lubridate_ymd_hms]								
	All	All	All	All	All	All	All	All	All
1	01-701-1015	1	<BLQ	<BLQ	0	2014-01-01T23:30:00	-1	Pre-dose	-0.5
2	01-701-1015	2	<BLQ	<BLQ		2014-01-02T00:05:00	1	5 Min Post-dose	0.08
3	01-701-1015	3	<BLQ	<BLQ		2014-01-02T00:30:00	1	30 Min Post-dose	0.5
4	01-701-1015	4	<BLQ	<BLQ		2014-01-02T01:00:00	1	1h Post-dose	1
5	01-701-1015	5	<BLQ	<BLQ		2014-01-02T01:30:00	1	1.5h Post-dose	1.5
6	01-701-1015	6	<BLQ	<BLQ		2014-01-02T02:00:00	1	2h Post-dose	2
7	01-701-1015	7	<BLQ	<BLQ		2014-01-02T04:00:00	1	4h Post-dose	4
8	01-701-1015	8	<BLQ	<BLQ		2014-01-02T06:00:00	1	6h Post-dose	6
9	01-701-1015	9	<BLQ	<BLQ		2014-01-02T08:00:00	1	8h Post-dose	8
10	01-701-1015	10	<BLQ	<BLQ		2014-01-02T12:00:00	1	12h Post-dose	12

Showing 1 to 10 of 500 entries

Previous

1

2

3

4

5

...

50

Next

The source viewer tab within **rock** is a utility designed to give quick access to view and explore source data files added to the environment. This capability is highly beneficial for data programmers and analysts, as it allows them to rapidly familiarize themselves with the content, structure, and quality of the source data. A clear understanding of the source data structure is necessary before documenting how variables will be derived. Since the source files may change over the course of a project (e.g., replacement of older files with updated ones), the source viewer allows user to ensure that they explore the current and correct version of the data.

THE SOURCE DATA ASSIGNMENT TO STUDY

VariablesSourceSource viewerDAT vs DDTLaTeX macrosGlobal infoExportHelpPreferences

Save AsSave

CDISCPIL0T01, ST1

It is recommended that studyOne, studyTwo etc are consistent with macros 'sone, 'stwo etc in project-macros.tex, see tab LaTeX macros

lstudyOneSTUDY1

lstudyTwoSTUDY2

lstudyThree

lstudyFour

Update file-to-study links and list of favorites

Replace source files (with newer versions)

Copy source column selection between source files

Clean study selection

Link files to the right study

Discard changes

Save changes

☐ Force re-read files?☐ Skip column summary?

Aliases should be formatted as: sourcefile=alias, where sourcefile matches exactly one of the files in the text area to the left. E.g. 20231231/ex.sas7bdat=alias1

New list of files for STUDY1 (lstudyOne). Must be only one file per line.

20250827/ST1/cgm.xpt
20250827/ST1/dm.sas7bdat
20250827/ST1/pc.sas7bdat
20250827/ST1/sv.xpt
20250827/ST1/vs.xpt
20250922/ST1/adex.sas7bdat
20250922/ST1/adlb.sas7bdat
20250922/ST1/adtte.sas7bdat
20250922/ST1/pc_collection.csv

New aliases for files for STUDY1 (lstudyOne). Must be only one alias per line.

New list of files for STUDY2 (lstudyTwo). Must be only one file per line.

20250827/ST2/cgm.xpt
20250827/ST2/dm.xpt
20250827/ST2/ex.xpt
20250827/ST2/lb.xpt
20250827/ST2/pc.xpt
20250827/ST2/sv.xpt
20250922/ST2/conmed.csv

New aliases for files for STUDY2 (lstudyTwo). Must be only one alias per line.

rock environment is designed to accommodate and manage different types of source data which can be imported and integrated in the framework. Managing source files efficiently is a core challenge in data programming, and **rock** addresses this by streamlining the handling of data updates. When updated source files are received, the system allows for near-instant replacement. This rapid replacement mechanism is a key factor in ensuring reproducibility, allowing the user to simply update the source data and re-execute the entire programming process. By standardizing how various source files are loaded and managed, **rock** ensures that the subsequent programming environment is properly configured to access the input data defined by the specifications.

SOURCE ASSIGNMENT TO VARIABLES

Source

Set Source to 'Produced'

Study	File name	Column	Specification	Derived from
STUDY1	20250827/ST1/pc.sas7bdat	PCSTRESN	where PCSPEC=PLASMA and PCTESTCD=XAN	☒
STUDY2	20250827/ST2/pc.xpt	PCSTRESN	where PCSPEC=PLASMA and PCTESTCD=X105	☒
				☒

The figure above visually represents the variable source editing interface which includes fields for study, file and variables, record selection, and derivation details. Once there is a clear understanding of which source data and columns are needed, the user can assign the specific source files and variables that contribute to its calculation, the necessary record selections applied, and some derivation details.

DATA SPECIFICATION TABLE PREVIEW



Data management report

Creation of DAT-1a-PMX-CONF-PMX-1.csv

Appendix 2. Data definition table

Table A2-1: Data definition table

Variable	Definition	Possible values	Description	Unit	Source
REF	Unique record number	Integer ≥ 1	Consecutive numbers starting from 1.	N/A	Produced
STUDYIDN	Study identifier	1, 2	1 = STUDY1 2 = STUDY2	N/A	STUDY1 Derived from dm.sas7bdat: STUDYID STUDY2 Derived from dm.xpt: STUDYID
ID	Unique subject identifier within current file	Integer ≥ 1	Consecutive numbers starting from 1 to max number of subjects within current file.	N/A	Produced
TYPE	Type of record	0, 1	0 = Dose event 1 = PK analyte plasma concentration	N/A	Produced
EVID	NONMEM event identification	0, 1	0 = Observation event 1 = Dose event	N/A	Produced
DV	NONMEM dependent variable	Numeric	Observation records (EVID=0): Observed value. BLQ=1: Set to the lower limit of quantification (LLOQ) divided by 2. Dose records (EVID=1): 0	ug/ml	STUDY1 Derived from pc.sas7bdat: PCSTRESN where PCSPEC=PLASMA and PCTESTCD=XAN STUDY2 Derived from pc.xpt: PCSTRESN where PCSPEC=PLASMA and PCTESTCD=X105
BLQ	Flag for concentration reported as below LLOQ in source data	0, 1	0 = No 1 = Yes	N/A	STUDY1 Derived from pc.sas7bdat: PCSTRESN, PCSTRESC STUDY2 Derived from pc.xpt: PCSTRESN, PCSTRESC Produced
LNDV	Log-transformed dependent variable	Numeric, DV	Natural logarithm of positive pharmacokinetic (PK) plasma observations. Rounded to 4 decimals. DV = Set to DV for other record	N/A	Produced

Following the detailed definition and linkage of all variables, **rock** generates a data specification table preview by one click, documenting the data requirement before the programming environment is implemented. The preview is automatically added later to the data programming report.

DATA PROGRAMMING REPORT GENERATION

Before the complete programming environment is exported, the final administrative details for the data programming report are collected. Through the **rock** interface, the user assigns the key project team members involved, such as the project leader, the data programmers responsible for the code, and the quality control reviewer to automatically enable their details in the data programming report for clear demonstration of accountability and roles. The data programming environment can then be exported.

Variables
Source
Source viewer
DAT vs DDT
LaTeX macros
Global info
Export
Help
Preferences

To import DDT into analysis report use latest REP template in PMX-RRW-PMX-1

DDT column widths, 6 numbers that sum up to 0.9, each > 0.02
0.08,0.13,0.12,0.21,0.03,0.33

Description breaks (Normally empty, see userguide)?

DMR template

QC reviewer?
Svetlana Freiberga
Create template DMR-1a-PMX-CONF-PMX-1
Preview code for R-RockSourceDataList.R
☒ use latest template in PMX-DMR-PMX-1
☐ use development template in PMX-DMR-PMX-1
☐ In code preview, include all source files linked to a study
☐ Overwrite SUB-OriginalData.Rnw when exporting DMR template

Preview of project-titlepage.tex in the shared LaTeX subfolder. Note that the title of the DMR is by default different from the project title. NB: preview not updated after project is opened in rock

```

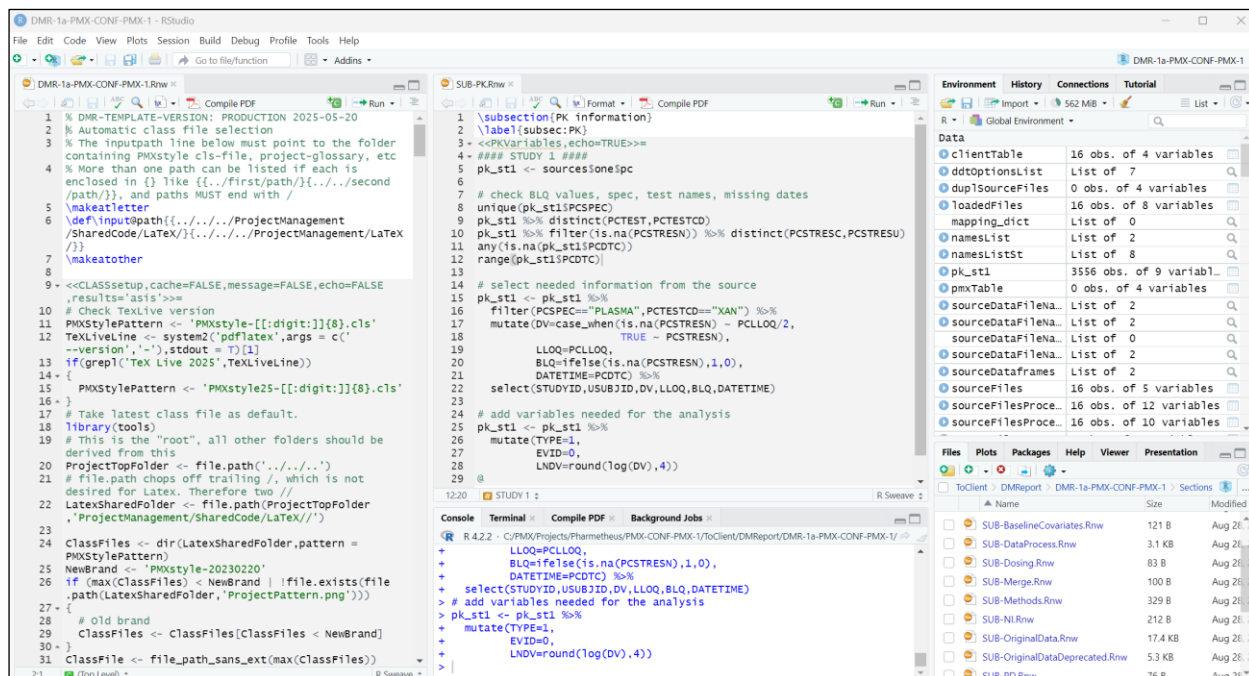
\projecttitle{\projectTitle}
\shortprojecttitle{\shortProjectTitle}
\reference{\pmxProjectName}
\confidential{\Confidential}
\pmtcontact{\pmxContact}
\emailpmtcontact{\pmxContactEmail}
\phonepmtcontact{+46 18 513 328}
\client{\longClientName}
\shortclient{\shortClientName}
\clientcontact{\clientContact}
\phoneclientcontact{\phoneClientContact}
\emailclientcontact{\emailClientContact}

```

PMX project leader?
DMR Author A?
DMR Author B
DMR Author C

Arnaud Gaby
Mounika Suddapalli
none
none

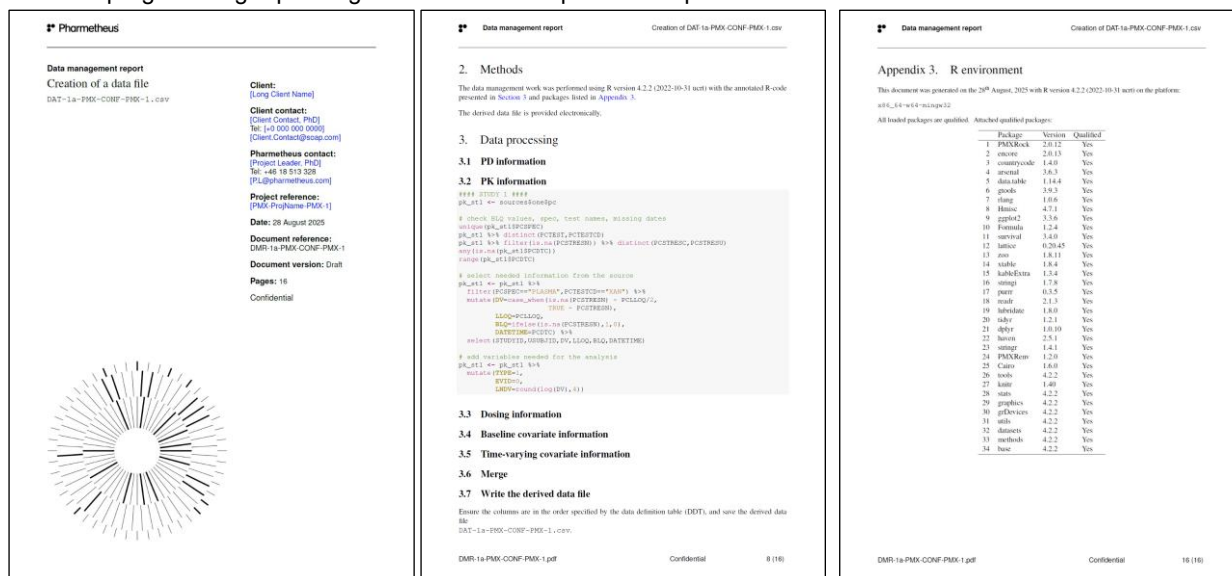
PROGRAMMING ENVIRONMENT FRAMEWORK



The framework features a standard directory structure alongside necessary setup files that properly load R libraries and correctly access the specified source data. To support programming work, it includes R script placeholders for specific analytical domains, such as dosing, PK, PD, and baseline covariates. This standardized framework allows documentation to be seamlessly integrated with the programming code and provides a structure which facilitates easier handover between team members.

DATA PROGRAMMING REPORT

The data programming report is generated as a comprehensive pdf file.



The core content of this report includes the executed programming code used to derive the deliverables, complete with all necessary comments and documentation. It automatically incorporates the finalized data specification table preview, to verify that documented requirements align with the delivered code.

AUTOMATED SANITY CHECKS

To ensure the accuracy and integrity of generated deliverables, **rock** incorporates a suite of automated sanity checks that support validation of datasets and specifications on multiple levels. It produces summary statistics for a quick overview of key data characteristics and identifies potential inconsistencies. Output dataset is continually compared against predefined specifications, verifying that all required variables, formats, and structures are correctly implemented.

VariablesSourceSource viewerDAT vs DDTLaTeX macrosGlobal infoExportHelpPreferences

csv-files in ProducedData/Dataset

DAT-1a-PMX-CONF-PMX-1.csv

compare file to DDT

auto-QC of DAT

Constants to use in auto-QC

'LLOQ'=1,'ULOQ'=1

Variable name of unique patient identifier
(e.g. ID or USUBJID)

ID

☐ Add continuity check to
'compare file to DDT'?

There are no columns with empty values.

Variable	Possible Values	Values in DAT	Classification
REF	Integer >= 1	min 1 median 3556.5 max 7112	numeric
STUDYIDN	1, 2	1, 2	numeric
ID	Integer >= 1	min 1 median 254.5 max 508	numeric
TYPE	0, 1	1	all_values_identical
EVID	0, 1	0	all_values_identical
DV	Numeric	min 0 median 0.0133377195463065 max 1.93809465994273	numeric
BLQ	0, 1	0, 1	numeric
LNDV	Numeric, DV	min -Inf median -4.31715 max 0.6617	numeric
LLOQ	Numeric >= 0	0.01	all_values_identical
SEXN	1, 2	1, 2	numeric
RACEN	1, 2, 3, 4, 5, 6, 7, -99	(4 unique values): 5, .., 3, 1	character
AGEYBL	Numeric > 0, -99	(37 unique values): 63, 64, 71, 74, 77, 85, 68, 81, 84, 52, 75, 57, 79, 56, ... (23 more unique values)	character
USUBJID	Character	(508 unique values): 01-701-1015, 01-701-1023, 01-701-1028, 01-701-1033, 01-701-1034, 01-701-1047, 01-701-1097, 01-701-1111, 01-701-1115, 01-701-1118, 01-701-1130, 01-701-1133, 01-701-1146, 01-701-1148, ... (494 more unique values)	character
DATETIME	Character	lubridate_ymd_hms	lubridate_ymd_hms

CONCLUSION

Using **rock** as the foundation for creating deliverables provides a robust framework that eases the project lifecycle for data programmers, quality controllers as well as modelers. Its standardized structure simplifies collaboration and handovers, while strict adherence to principles of reproducibility ensures that deliverables can be reliably regenerated from updated source data. Furthermore, built-in version control contributes to full traceability, supporting transparency and accountability throughout the project development. **rock**'s extensible design allows teams to incorporate additional specification templates as needed, and its flexibility in supporting diverse programming report formats makes it adaptable to a wide range of project requirements. Together, these advantages make **rock** a powerful solution for consistent, traceable, and high-quality deliverable creation.

CONTACT INFORMATION

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

Author Name: Svetlana Freiberga

Company: Pharmetheus

Address: Kungsängstull 4, 753 19 Uppsala, Sweden

Work Phone: [+46\(0\)18-51 33 28](tel:+46018513328)

Email: svetlana.freiberga@pharmetheus.com

Website: <https://pharmetheus.com/>