Package: r4subdata (via r-universe)

May 15, 2026

Title Example Datasets for Clinical Submission Readiness

Version 0.1.2

Description Provides realistic synthetic example datasets for the R4SUB (R for Regulatory Submission) ecosystem. Includes a pharma study evidence table, ADaM (Analysis Data Model) and SDTM (Study Data Tabulation Model) metadata following CDISC (Clinical Data Interchange Standards Consortium) conventions (<`https://www.cdisc.org`>), traceability mappings, a risk register based on ICH (International Council for Harmonisation) Q9 quality risk management principles (<`https://www.ich.org/page/quality-guidelines`>), and regulatory indicator definitions. Designed for demos, vignettes, and package testing.

License MIT + file LICENSE

URL `https://r4sub.github.io/r4subdata/`, `https://github.com/R4SUB/r4subdata`

BugReports `https://github.com/R4SUB/r4subdata/issues`

Depends R (>= 4.2)

Imports tibble

Suggests knitr, rmarkdown, testthat (>= 3.0.0)

VignetteBuilder knitr

Config/testthat/edition 3

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.3.3

LazyData true

Repository https://r4sub.r-universe.dev

Date/Publication 2026-03-16 11:21:23 UTC

RemoteUrl https://github.com/r4sub/r4subdata

RemoteRef HEAD

RemoteSha 1164e7719256e6fffad69f637a6f58184fd4d04f

Contents

`adam_metadata` *ADaM Variable-Level Metadata*

Description

ADaM (Analysis Data Model) variable-level metadata for ADSL (Subject-Level Analysis Dataset, 16 vars), ADAE (Adverse Events Analysis Dataset, 10 vars), and ADLB (Laboratory Results Analysis Dataset, 10 vars). Follows CDISC (Clinical Data Interchange Standards Consortium) ADaM conventions.

Usage

```
adam_metadata
```

Format

A tibble with 36 rows and 6 columns:

dataset Character. ADaM dataset name (ADSL, ADAE, ADLB).

variable Character. Variable name.

label Character. Variable label.

type Character. Variable type (Char or Num).

length Integer. Variable length.

format Character. SAS (Statistical Analysis System) format (or NA).

Source

Synthetic metadata based on CDISC ADaM (Analysis Data Model) standards.

Examples

```
data(adam_metadata)
table(adam_metadata$dataset)
```

`dataset_dictionary` *Dataset Column Dictionary*

Description

Returns column names, types, and descriptions for a given r4subdata dataset.

Usage

```
dataset_dictionary(dataset)
```

Arguments

`dataset`	Character. Name of the dataset (e.g., `"evidence_pharma"`).

Value

A tibble with columns: `column`, `type`, `description`.

Examples

```
dataset_dictionary("evidence_pharma")
dataset_dictionary("adam_metadata")
```

`evidence_pharma` *Pharma Study Evidence Table*

Description

A realistic evidence table for study CDISCPILOT01 (Clinical Data Interchange Standards Consortium Pilot Study 01) covering all four R4SUB (R for Regulatory Submission) pillars (quality, trace, risk, usability) with 250 rows and 18 indicators across multiple datasets and sources.

Usage

```
evidence_pharma
```

Format

A tibble with 250 rows and 17 columns:

run_id Character. Unique run identifier.

study_id Character. Study identifier (CDISCPILOT01).

asset_type Character. Asset type: dataset, define, program, validation, spec, other.

asset_id Character. Asset identifier (e.g., ADSL, define.xml).

source_name Character. Source of the evidence (e.g., pinnacle21).

source_version Character. Version of the source tool.

indicator_id Character. Indicator identifier (e.g., Q-MISS-VAR).

indicator_name Character. Human-readable indicator name.

indicator_domain Character. Domain: quality, trace, risk, usability.

severity Character. Severity: info, low, medium, high, critical.

result Character. Result: pass, fail, warn, na.

metric_value Numeric. Metric value (if applicable).

metric_unit Character. Unit for metric_value.

message Character. Descriptive message.

location Character. Location reference (e.g., ADSL:AGE).

evidence_payload Character. JSON payload with additional details.

created_at POSIXct. Timestamp when evidence was created.

Source

Synthetic data based on the CDISC (Clinical Data Interchange Standards Consortium) Pilot Study 01 structure.

Examples

```
data(evidence_pharma)
head(evidence_pharma)
table(evidence_pharma$indicator_domain)
```

`list_datasets` *List Available r4subdata Datasets*

Description

Returns a summary of all datasets included in the r4subdata package.

Usage

```
list_datasets()
```

Value

A tibble with columns: `name`, `description`, `n_rows`, `n_cols`.

Examples

```
list_datasets()
```

`oncology_evidence` *Oncology Trial Evidence Table*

Description

A synthetic R4SUB evidence table for study ONCO-2025-001 covering all four R4SUB (R for Regulatory Submission) pillars (quality, trace, risk, usability) with 29 rows across ADSL, ADRS, and ADTTE datasets. Demonstrates realistic evidence patterns for an oncology submission with mixed pass/warn/fail results.

Usage

```
oncology_evidence
```

Format

A tibble with 29 rows and 17 columns:

run_id	Character. Unique run identifier.
study_id	Character. Study identifier (ONCO-2025-001).
asset_type	Character. Asset type: dataset, define, program, validation, spec, other.
asset_id	Character. Asset identifier (e.g., ADSL, ADRS).
source_name	Character. Source tool name.
source_version	Character. Version of the source tool.
indicator_id	Character. Indicator identifier (e.g., Q-MISS-VAR).
indicator_name	Character. Human-readable indicator name.
indicator_domain	Character. Domain: quality, trace, risk, usability.
severity	Character. Severity: info, low, medium, high, critical.
result	Character. Result: pass, fail, warn, na.
metric_value	Numeric. Metric value (if applicable).
metric_unit	Character. Unit for metric_value.
message	Character. Descriptive message.
location	Character. Location reference (e.g., ADRS:AVAL).
evidence_payload	Character. JSON payload with additional details.
created_at	POSIXct. Timestamp when evidence was created.

Source

Synthetic evidence data for a Phase II oncology trial.

Examples

```
data(oncology_evidence)
table(oncology_evidence$indicator_domain)
table(oncology_evidence$result)
```

`oncology_metadata` *Oncology Trial Variable-Level Metadata*

Description

ADaM (Analysis Data Model) variable-level metadata for a synthetic oncology trial covering ADSL (Subject-Level Analysis Dataset, 14 vars), ADRS (Response Analysis Dataset, 10 vars), and ADTTE (Time-to-Event Analysis Dataset, 8 vars). Includes origin, derivation, and codelist columns suitable for use with `r4subusability` assessments.

Usage

```
oncology_metadata
```

Format

A tibble with 32 rows and 7 columns:

dataset Character. ADaM dataset name (ADSL, ADRS, ADTTE).

variable Character. Variable name.

label Character. Variable label.

origin Character. Variable origin (CRF, Derived, Assigned).

derivation Character. Derivation text (NA if not derived).

codelist Character. CDISC (Clinical Data Interchange Standards Consortium) codelist code (NA if not applicable).

type Character. Variable type (Char or Num).

Source

Synthetic metadata for a Phase II oncology trial following CDISC (Clinical Data Interchange Standards Consortium) ADaM conventions.

Examples

```
data(oncology_metadata)
table(oncology_metadata$dataset)
table(oncology_metadata$origin)
```

`regulatory_indicators` *Regulatory Indicator Definitions*

Description

Reference table of 30 indicator definitions across all four R4SUB (R for Regulatory Submission) domains (quality, trace, risk, usability). Each indicator has a unique ID, default severity, typical source, and descriptive tags.

Usage

```
regulatory_indicators
```

Format

A tibble with 30 rows and 7 columns:

indicator_id Character. Unique indicator identifier.

indicator_name Character. Human-readable indicator name.

domain Character. Indicator domain: quality, trace, risk, usability.

description Character. Detailed description.

severity_default Character. Default severity level.

source Character. Typical source tool.

tags Character. Comma-separated tags.

Source

Curated indicator definitions for the R4SUB (R for Regulatory Submission) ecosystem.

Examples

```
data(regulatory_indicators)
table(regulatory_indicators$domain)
```

`risk_register_pharma` *Pharma Risk Register*

Description

A Failure Mode and Effects Analysis (FMEA)-based risk register with 18 risks covering data quality, traceability, documentation, programming, and compliance categories. Includes probability, impact, and detectability scores on a 1-5 scale. Structured according to ICH (International Council for Harmonisation) Q9 quality risk management principles.

Usage

```
risk_register_pharma
```

Format

A tibble with 18 rows and 9 columns:

risk_id Character. Unique risk identifier (RISK-001 to RISK-018).

description Character. Risk description.

category Character. Risk category.

probability Integer. Probability of occurrence (1-5).

impact Integer. Impact severity (1-5).

detectability Integer. Detectability rating (1-5).

owner Character. Risk owner name.

mitigation Character. Mitigation action (or NA).

status Character. Status: open, mitigated, closed, accepted.

Source

Synthetic risk register based on ICH (International Council for Harmonisation) Q9 quality risk management principles.

Examples

```
data(risk_register_pharma)
table(risk_register_pharma$category)
```

`sdtm_metadata` *SDTM Variable-Level Metadata*

Description

SDTM (Study Data Tabulation Model) variable-level metadata for DM (Demographics, 17 vars), AE (Adverse Events, 14 vars), and LB (Laboratory Results, 12 vars). Follows CDISC (Clinical Data Interchange Standards Consortium) SDTM conventions.

Usage

```
sdtm_metadata
```

Format

A tibble with 43 rows and 6 columns:

dataset Character. SDTM domain name (DM, AE, LB).

variable Character. Variable name.

label Character. Variable label.

type Character. Variable type (Char or Num).

length Integer. Variable length.

format Character. SAS (Statistical Analysis System) format (or NA).

Source

Synthetic metadata based on CDISC SDTM (Study Data Tabulation Model) standards.

Examples

```
data(sdtm_metadata)
table(sdtm_metadata$dataset)
```

`trace_mapping` *ADaM-to-SDTM Traceability Mapping*

Description

Maps ADaM (Analysis Data Model) variables to their SDTM (Study Data Tabulation Model) source variables with derivation text and confidence scores. Includes direct copies, derived variables, and unmapped entries. Follows CDISC (Clinical Data Interchange Standards Consortium) traceability conventions.

Usage

```
trace_mapping
```

Format

A tibble with 25 rows and 6 columns:

adam_dataset Character. Source ADaM dataset.

adam_var Character. Source ADaM variable.

sdtm_domain Character. Target SDTM domain (NA if derived).

sdtm_var Character. Target SDTM variable (NA if derived).

derivation_text Character. Derivation description text.

confidence Numeric. Mapping confidence score (0-1, NA if unmapped).

Source

Synthetic traceability mapping based on CDISC conventions.

Examples

```
data(trace_mapping)
table(trace_mapping$adam_dataset)
```

Index