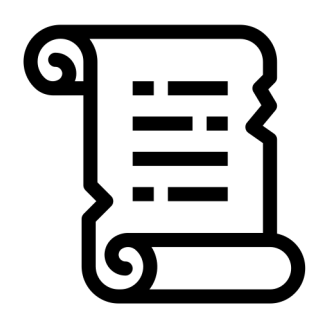
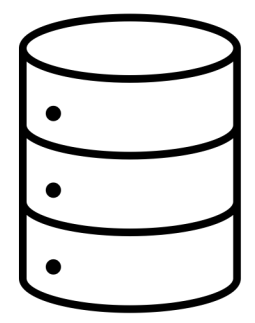


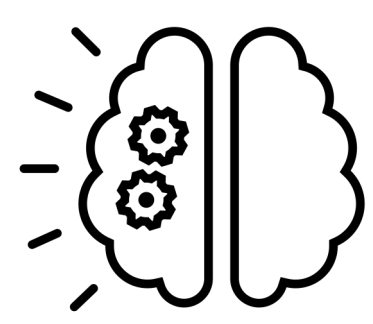
### Introduction



Historical clinical trials contain a wealth of information that can be used for insight generation within life-sciences organisations.



Good quality, interoperable data from historical trials fills the gap between high-quality, small-scale pooling on the one hand, and huge datasets full of rather messy real-world data on the other.



It can be used for applications, including data for synthetic control arms, identification of patients of interest for specific disease areas, and for the development and training of AI models.

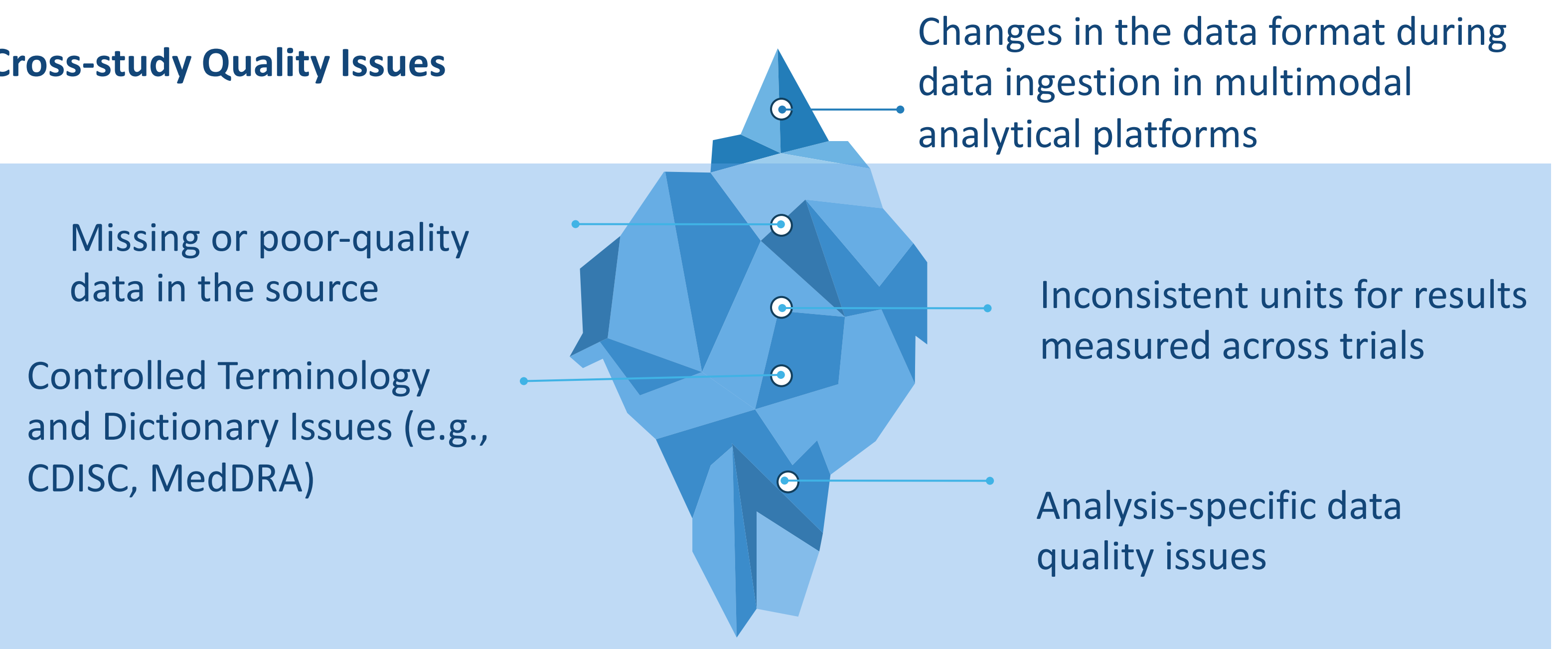


There are several challenges to using the data and there is no clear framework to measure the quality of the data.



Submission-oriented checks like Pinnacle21 focus on compliance within a study, and don't address cross-study quality issues.

#### Cross-study Quality Issues



### Solution



#### Align Data Standards

We chose the Study Data Tabulation Model (SDTM) described by CDISC for cross-study pooling and analysis of data as it allows us to pool the data without prior knowledge of the analysis needs



#### Measure Data Quality

We devised quality metrics that assess aspects such as:

- Cross domain consistency of the data
- Alignment of the content of the data against code lists and dictionaries (CDISC, MedDRA, etc.)
- Format of the data (ISO Dates, etc.)

These metrics allow experts to govern and improve the overall quality of the data

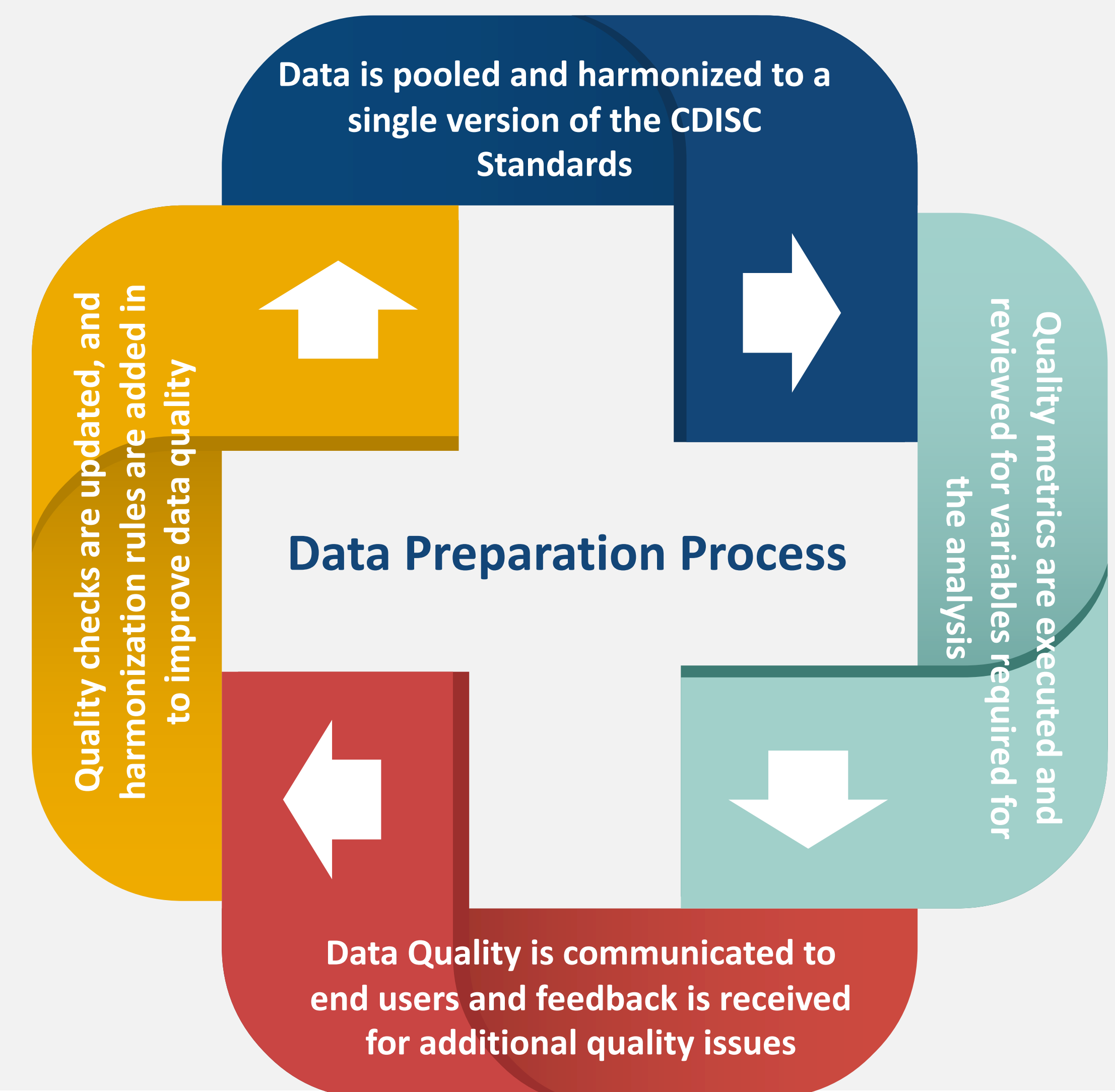
#### Implement Standards

We standardised all the studies to a single version of the SDTM IG, a single version of the code lists for the controlled terminology, and a single version of dictionaries (e.g., MedDRA)

#### Improve Data Quality

We store the history of the execution of our quality metrics so that Data Engineers can compare and quantify the improvement in the quality of the data over time

### Iterative process to provide data for secondary analysis



### Results

1

#### Global overview of metrics that can be leveraged by experts to monitor and improve data quality

FOUNDATA Harmonisation Metrics for SDTM  
Global overview

Visualisation: [Full metrics list](#) Study: [ALL STUDIES](#) Timestamp: 2024-04-24T17:36:00

Metric generation started on Thursday, 4 April 2024 at 17:36 (runtime 21 minutes)

Category	Metric Name	Value	Trend
AE	AE0001	100%	▲
	AE0002	100%	▲
	AE0003	100%	▲
	AE0004	100%	▲
	AE0005	100%	▲
	AE0006	100%	▲
	AE0007	100%	▲
	AE0008	100%	▲
	AE0009	100%	▲
	AE0010	100%	▲
CM	CM0001	100%	▲
	CM0002	100%	▲
	CM0003	100%	▲
	CM0004	100%	▲
	CM0005	100%	▲
	CM0006	100%	▲
	CM0007	100%	▲
	CM0008	100%	▲
	CM0009	100%	▲
	CM0010	100%	▲
DM	DM0001	100%	▲
	DM0002	100%	▲
	DM0003	100%	▲
	DM0004	100%	▲
	DM0005	100%	▲
	DM0006	100%	▲
	DM0007	100%	▲
	DM0008	100%	▲
	DM0009	100%	▲
	DM0010	100%	▲
DS	DS0001	100%	▲
	DS0002	100%	▲
	DS0003	100%	▲
	DS0004	100%	▲
	DS0005	100%	▲
	DS0006	100%	▲
	DS0007	100%	▲
	DS0008	100%	▲
	DS0009	100%	▲
	DS0010	100%	▲
EK	EK0001	100%	▲
	EK0002	100%	▲
	EK0003	100%	▲
	EK0004	100%	▲
	EK0005	100%	▲
	EK0006	100%	▲
	EK0007	100%	▲
	EK0008	100%	▲
	EK0009	100%	▲
	EK0010	100%	▲
LB	LB0001	100%	▲
	LB0002	100%	▲
	LB0003	100%	▲
	LB0004	100%	▲
	LB0005	100%	▲
	LB0006	100%	▲
	LB0007	100%	▲
	LB0008	100%	▲
	LB0009	100%	▲
	LB0010	100%	▲
MH	MH0001	100%	▲
	MH0002	100%	▲
	MH0003	100%	▲
	MH0004	100%	▲
	MH0005	100%	▲
	MH0006	100%	▲
	MH0007	100%	▲
	MH0008	100%	▲
	MH0009	100%	▲
	MH0010	100%	▲
VS	VS0001	100%	▲
	VS0002	100%	▲
	VS0003	100%	▲
	VS0004	100%	▲
	VS0005	100%	▲
	VS0006	100%	▲
	VS0007	100%	▲
	VS0008	100%	▲
	VS0009	100%	▲
	VS0010	100%	▲

2

#### Example queries that are executed to produce quality metrics

FOUNDATA Harmonisation Metrics for SDTM  
Full metrics list

Visualisation: [Full metrics list](#) Study: [ALL STUDIES](#) Timestamp: 2024-04-24T17:36:00

Metric generation started on Thursday, 4 April 2024 at 17:36 (runtime 21 minutes)

AE CM DM DS EK LB MH VS

Category	Metric Name	Value	Trend
CM - Concomitant/Prior Medications	CMCLAS - Medication Class	68.6%	+59.8% ▲
	CMCLASCD - Medication Class Code	68.6%	+68.6% ▲
	CMDECCO - Standardized Medication Name	100%	+7.2% ▲
MH - Medical History	DOMAIN - Domain Abbreviation	72.6%	+42.9% ▲
	MHHLGT - High Level Group Term	100%	+84.9% ▲
MHBLT - Body System or Organ Class Code	MHBLT - High Level Term	100%	+73.2% ▲
	MHLLT - Lowest Level Term	100%	+73.2% ▲

Supplementing technical checks on data quality with functional checks helps ensure clinical trial data is “analysis-ready” for secondary purposes. CDISC standards provide a perfect framework for implementation of these checks.