Aim for the stars - the use of metadata to manage your clinical data

PhUSE 2017 SDE - Chilly-Mazarin, France

PRESENTER / Niels Both

DATE / 12.september 2017

AGENDA – Aim for the stars

CDISC and Metadata

Why is the flow of clinical data so complicated?

Metadata to help the dataflow and definition of datapoints

Metadata structure & technology

Ideas for successful MDR use

S-CUBED

Niels Both / S-CUBED

Niels Both
nb@s-cubed.dk

CDISC Expert
SAS Programmer

20 Years experience

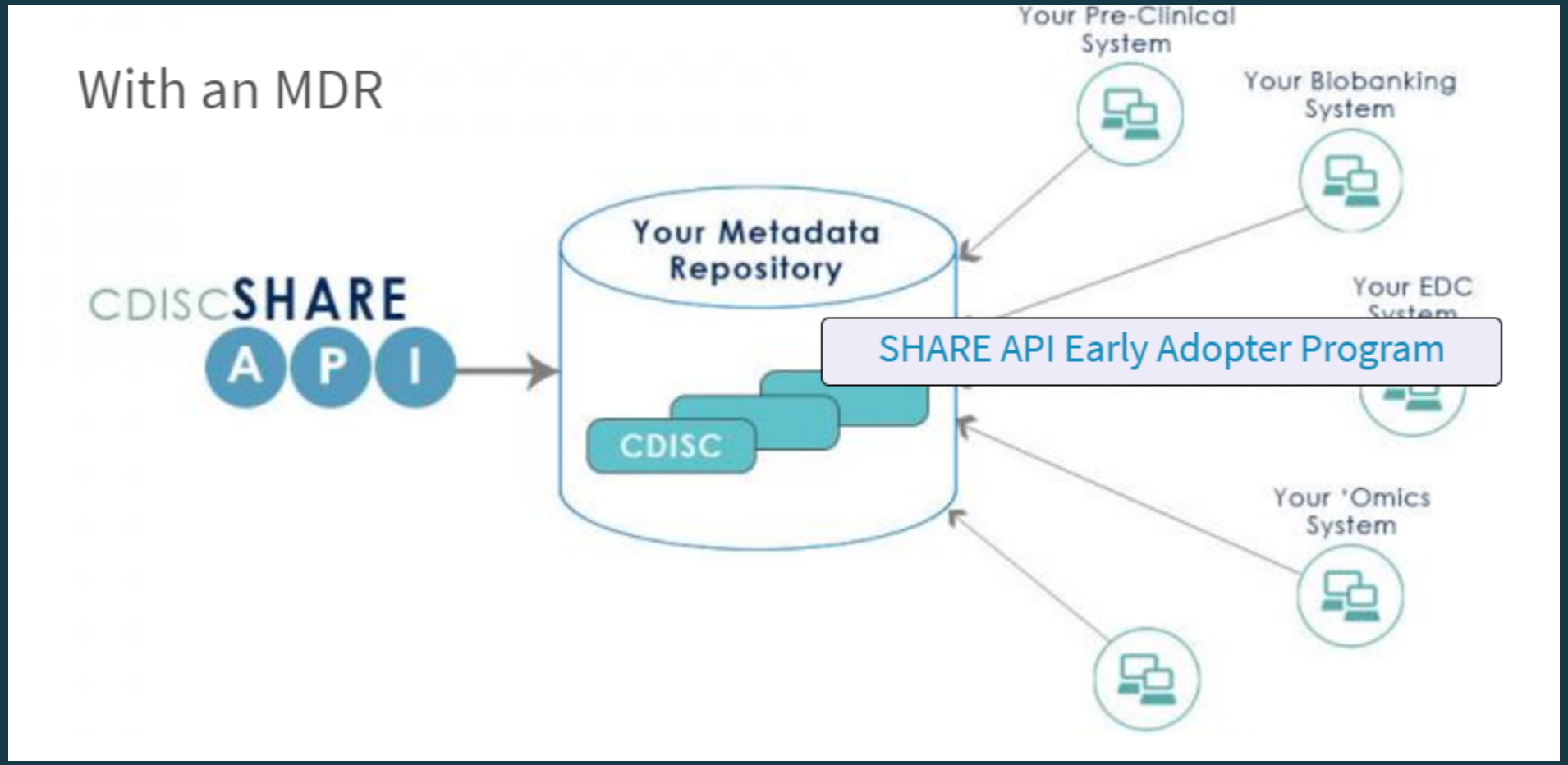
S-CUBED
Consultancy with offices in Denmark and UK
Specialized to assist the industry with all aspects of handling of clinical data and information

S-CUBED

CDISC and Metadata

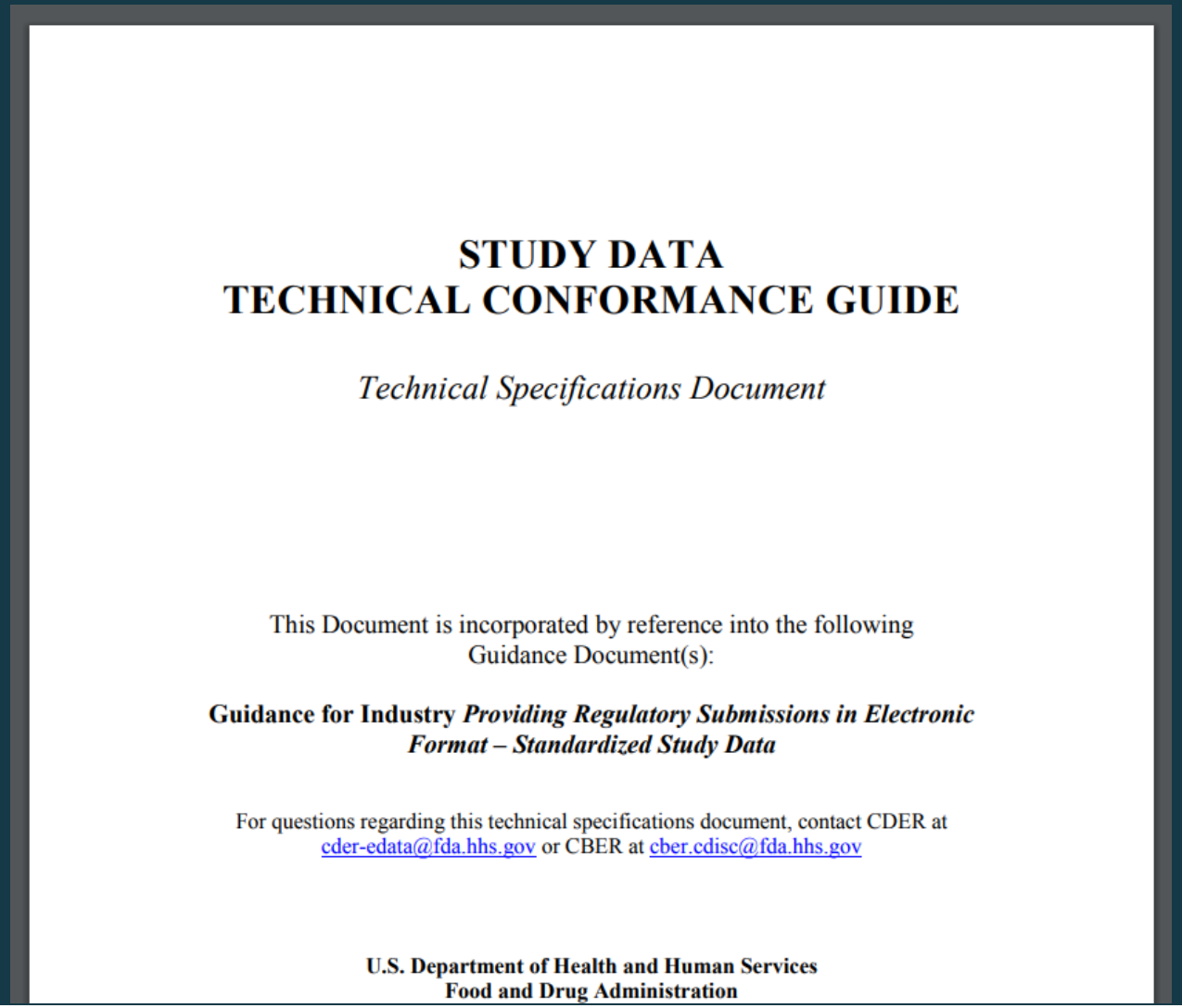
S-CUBED

CDISC and Metadata



S-CUBED

CDISC and Metadata



STUDY DATA
TECHNICAL CONFORMANCE GUIDE

Technical Specifications Document

This Document is incorporated by reference into the following Guidance Document(s):

Guidance for Industry *Providing Regulatory Submissions in Electronic Format – Standardized Study Data*

For questions regarding this technical specifications document, contact CDER at cder-edata@fda.hhs.gov or CBER at cber.cdisc@fda.hhs.gov

U.S. Department of Health and Human Services
Food and Drug Administration

S-CUBED

CDISC and Metadata

Contains Nonbinding Recommendations

FDA U.S. FOOD & DRUG ADMINISTRATION

Appendix: Data Standards and Interoperable Data Exchange

This appendix provides some of the guiding principles for the Agency's long-term study data standards management strategies. An important goal of standardizing study data submissions is to achieve an acceptable degree of *semantic interoperability* (discussed below). This appendix describes different types of interoperability and how data standards can support interoperable data exchange now and in the future.

At the most fundamental level, study data can be considered a collection of data elements and their relationships. A data element is the smallest (or *atomic*) piece of information that is useful for analysis (e.g., a systolic blood pressure measurement, a lab test result, a response to a question on a questionnaire).

A data value is by itself meaningless without additional information about the data (so called *metadata*). Metadata is often described as *data about data*. Metadata is structured information that describes, explains, or otherwise makes it easier to retrieve, use, or manage data.[53] For example, the number *44* itself is meaningless without an association with Hematocrit and the unit of measurement (e.g. "%"). Hematocrit in this example is metadata that further describes the data.

Just as it is important to standardize the representation of data (e.g., M and F for male and female, respectively), it is equally important to standardize the metadata. The expressions Hematocrit = 44; Hct = 44, or Hct Lab Test = 44 all convey the same information to a human, but an information system or analysis program will fail to recognize that they are equivalent because the metadata is not standardized. It is also important to standardize the definition of the metadata, so that the meaning of a Hematocrit value is constant across studies and submissions.

In addition to standardizing the data and metadata, it is important to capture and represent relationships (also called associations) between data elements in a standard way.

S-CUBED

Why is the flow of clinical data so complicated ?

S-CUBED

Why is the flow of clinical data so complicated ?

S-CUBED

Why is the flow of clinical data so complicated ?

S-CUBED

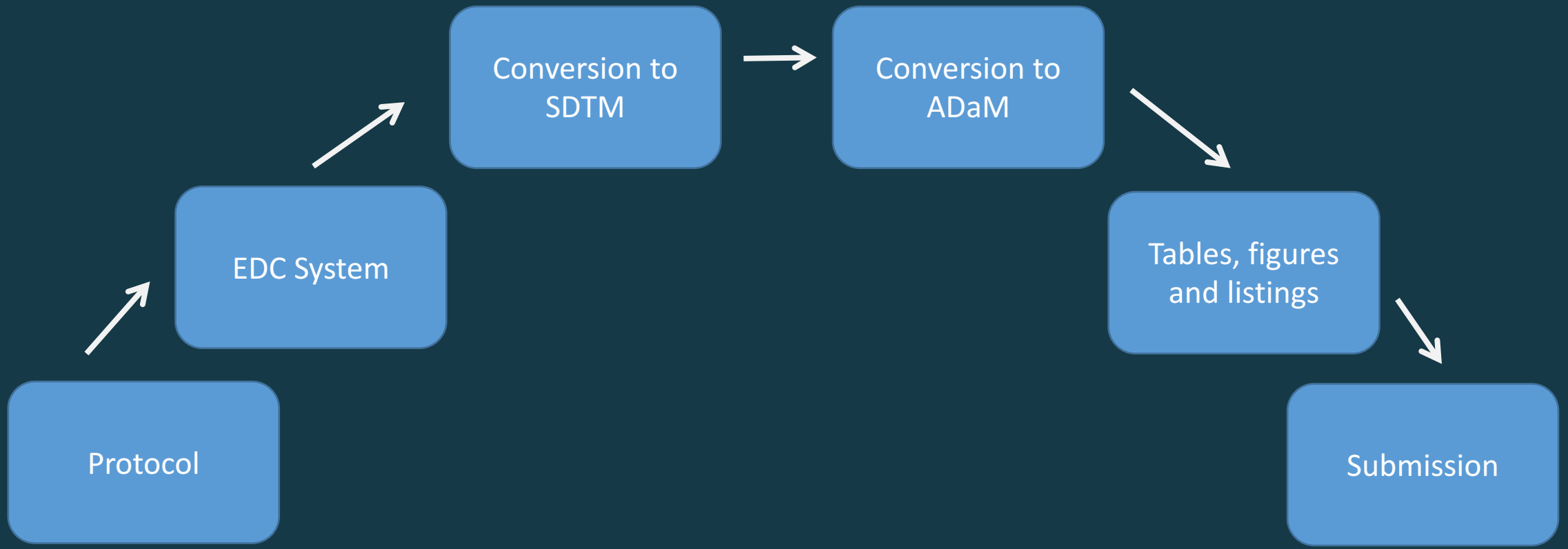
Why is the flow of clinical data so complicated ?

S-CUBED

Why is the flow of clinical data so complicated ?

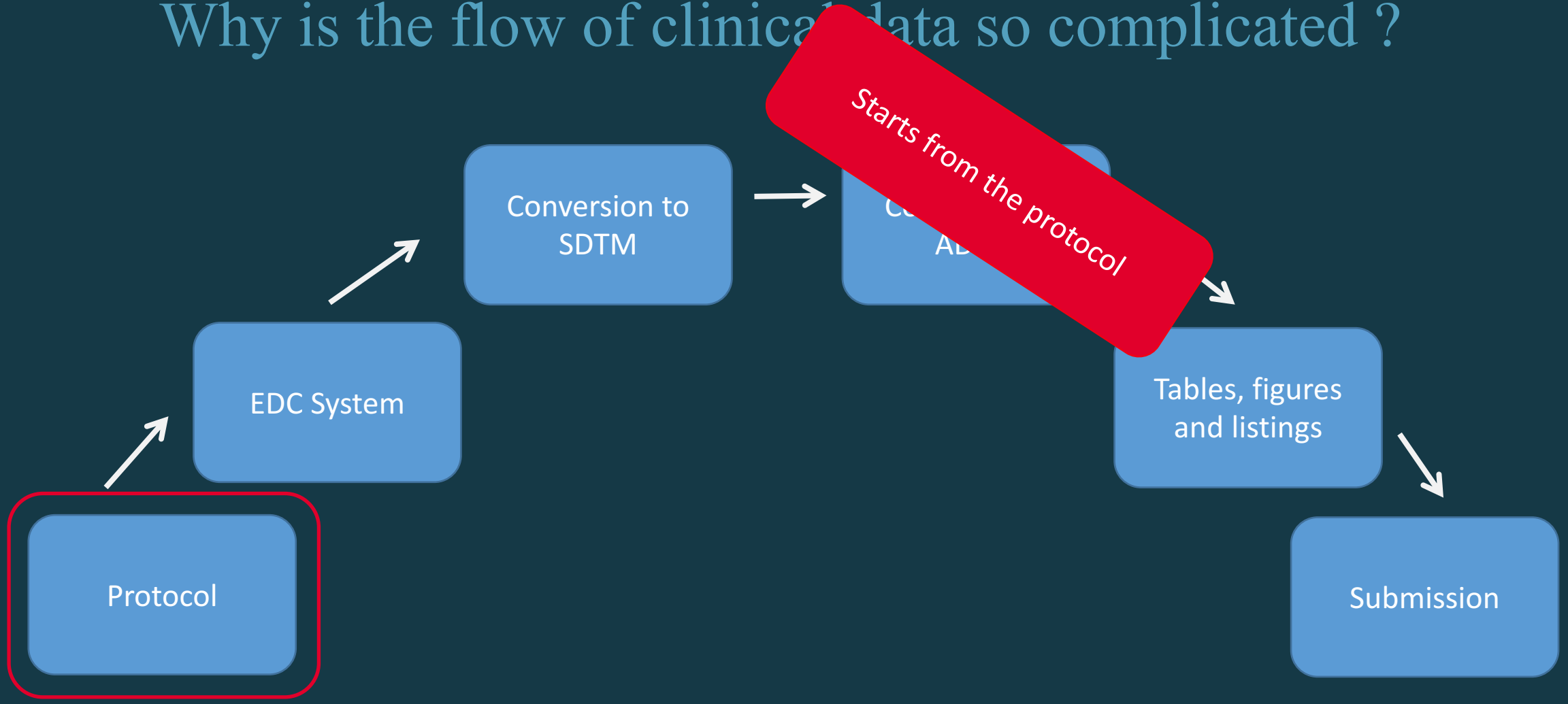
S-CUBED

Why is the flow of clinical data so complicated ?



S-CUBED

Why is the flow of clinical data so complicated ?



S-CUBED

The protocol and its many stakeholders

The protocol and its many stakeholders

MW: It must be written in a clear precise and understandable way

S-CUBED

The protocol and its many stakeholders

MW: It must be written in a clear precise and understandable way

Statistician: It is needed to collect the data in this way in order to get a significant result

The protocol and its many stakeholders

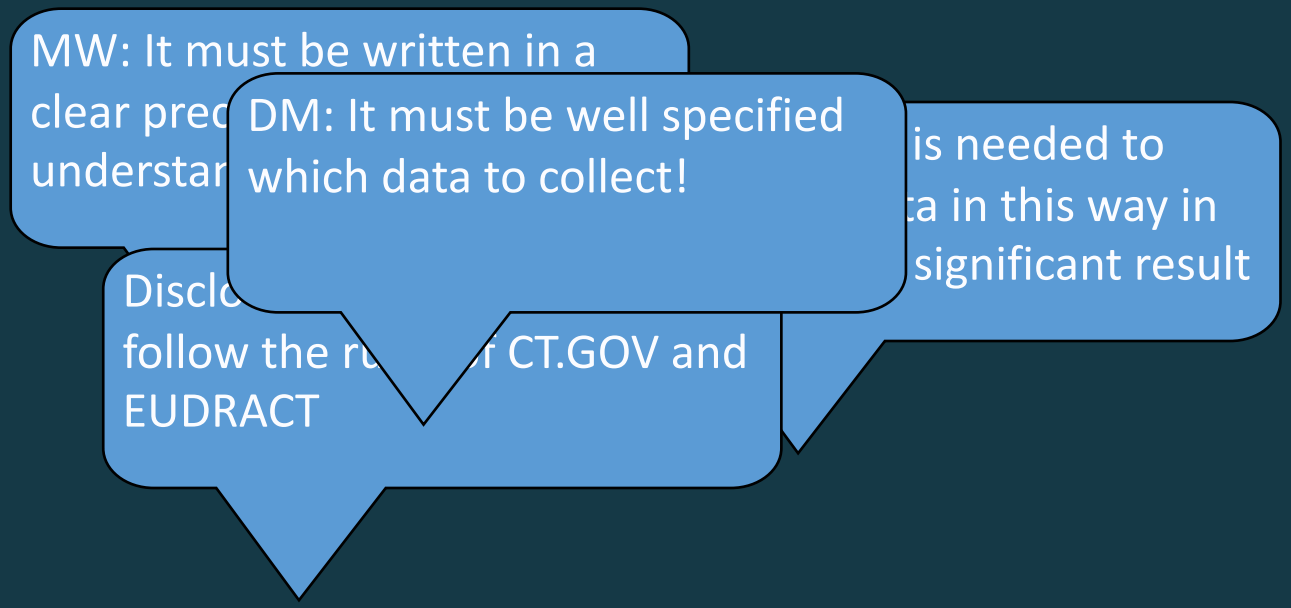
MW: It must be written in a clear precise and understandable way

Statistician: It is needed to collect the data in this way in [illegible] to get a significant result

Disclosure people: We must follow the rules of CT.GOV and EUDRACT

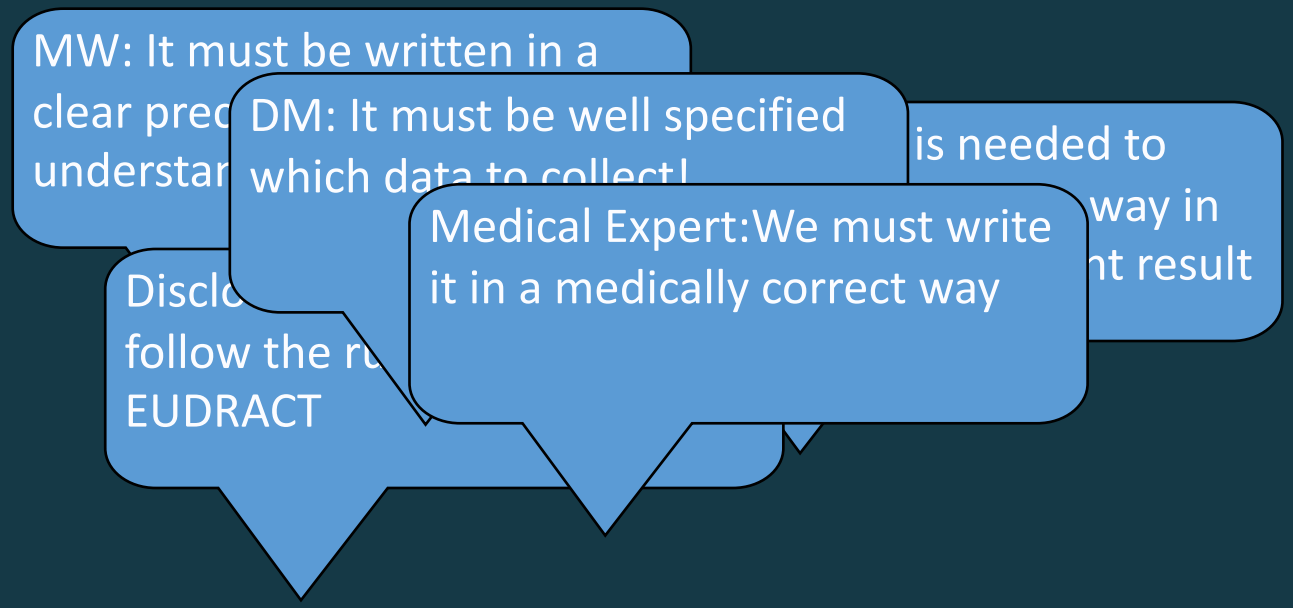
S-CUBED

The protocol and its many stakeholders



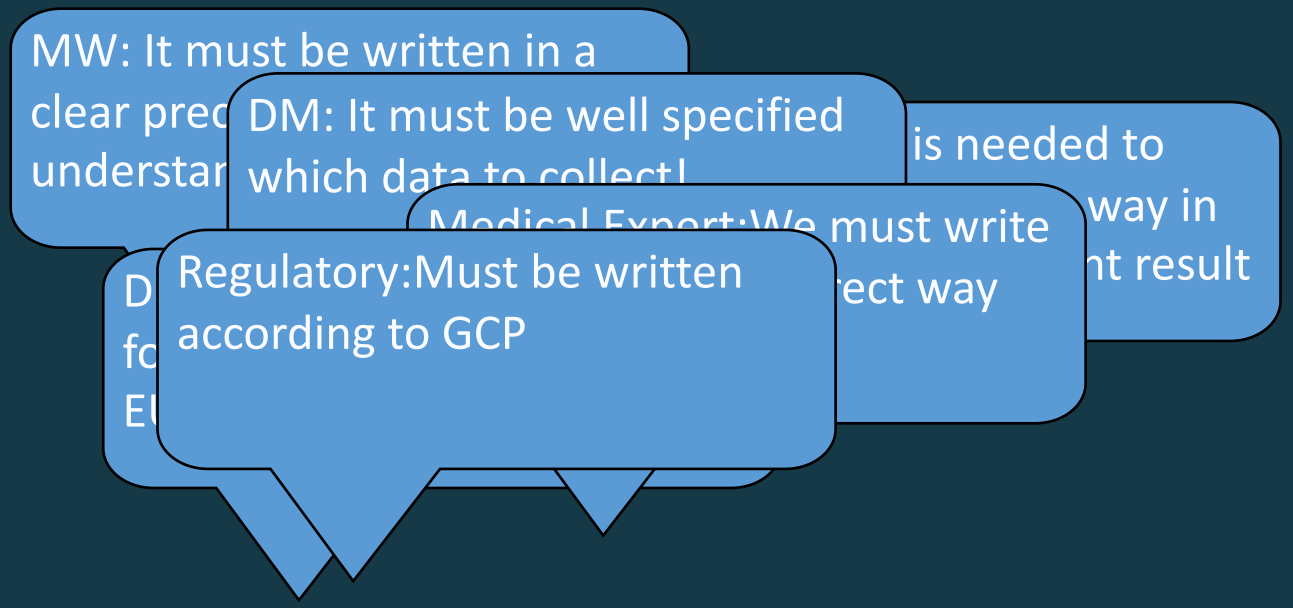
S-CUBED

The protocol and its many stakeholders

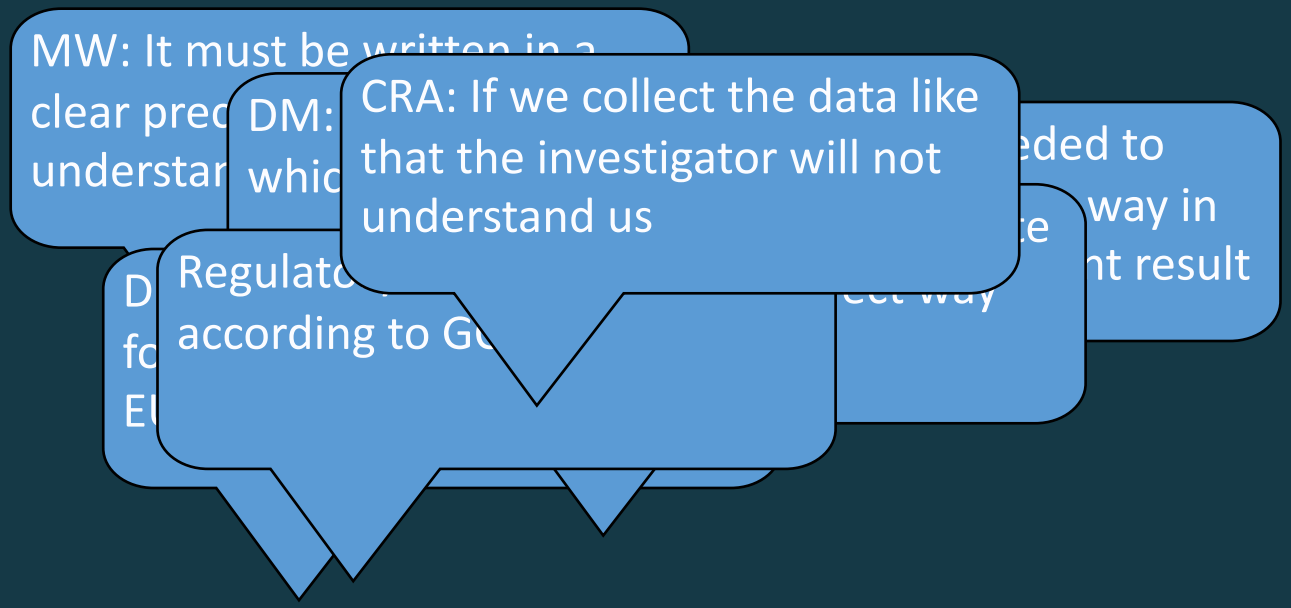


S-CUBED

The protocol and its many stakeholders

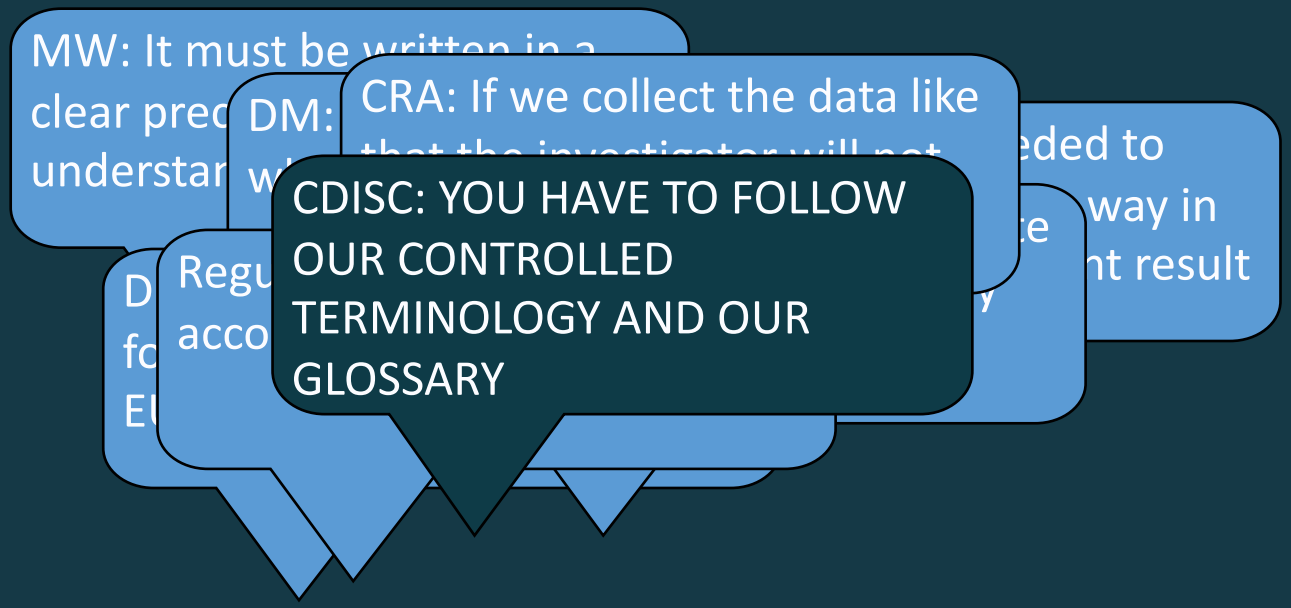


The protocol and its many stakeholders



S-CUBED

The protocol and its many stakeholders



S-CUBED

The protocol and its many stakeholders

MW: It must be written in a clear prec understar

DM: w

CRA: If we collect the data like

eded to way in nt result

Regu acco

CDISC: YOU HAVE TO FOLLOW OUR CONTROLLED TERMINOLOGY AND OUR GLOSSARY

S-CUBED

The protocol and its many stakeholders

S-CUBED

The grim reality……. PROCESS

In many pharmaceutical companies:

- The implementation of CDISC standards happens further down the dataflow than it ought to.
- In the worst case - CDISC compliance problems are fixed just prior to submission by a CRO or an SDTM programmer.
- In the worst case - Information about what was fixed is not necessarily fed back to the organisation as input to the protocols for the next trials.

S-CUBED

The grim reality…….
STANDARDS

- Standards are dynamic. New controlled terminology is issued several times a year.

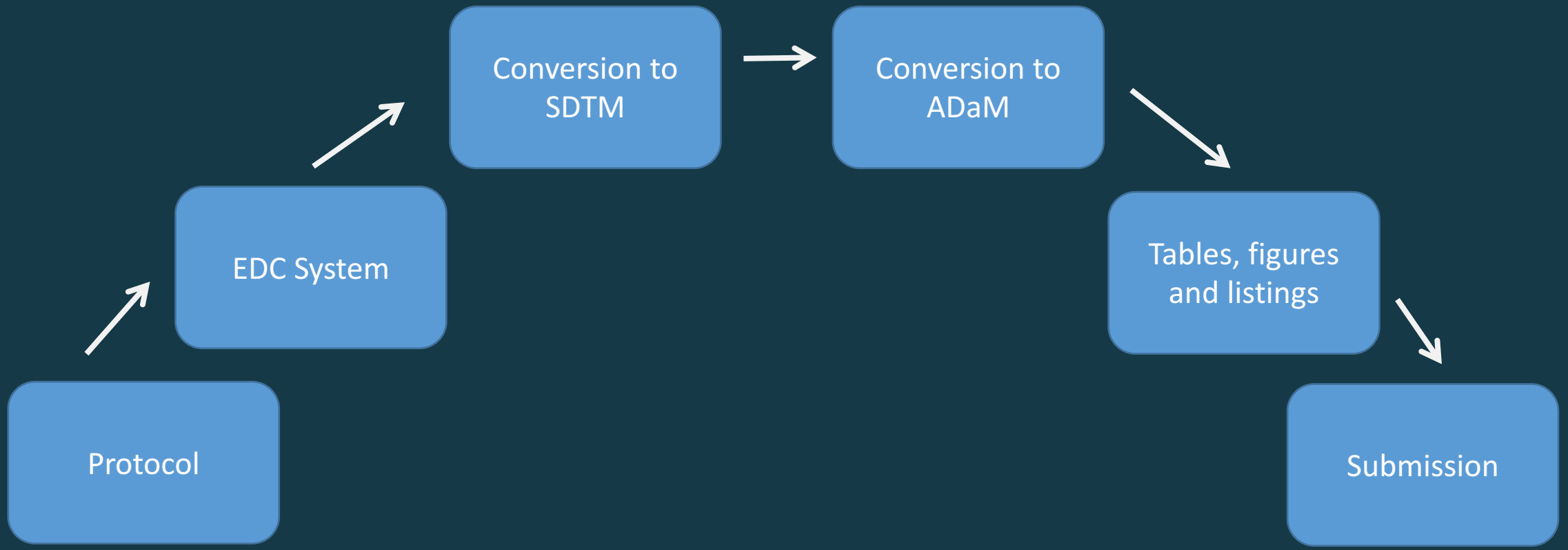
- From FDA: New (very) detailed requirements about how to submit data are issued a few times per year. In some cases the requirements from FDA contradict the formal rules of the standards.

- The CDISC standards are not flawless, so there can be inconsistencies or lack of clarity that that the organisation and the computer systems must be ready to cope with.

- Different regulatory bodies have different ideas and different definitions.

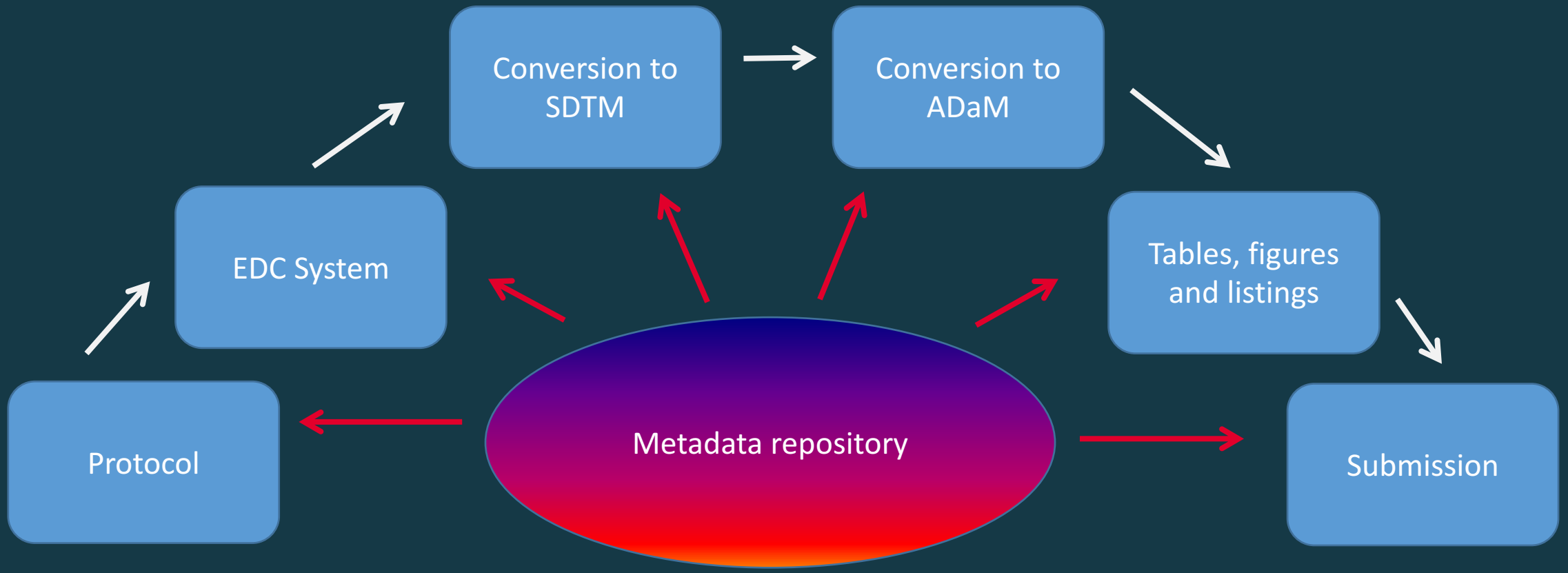
S-CUBED

An MDR controls definitions and versions of standards



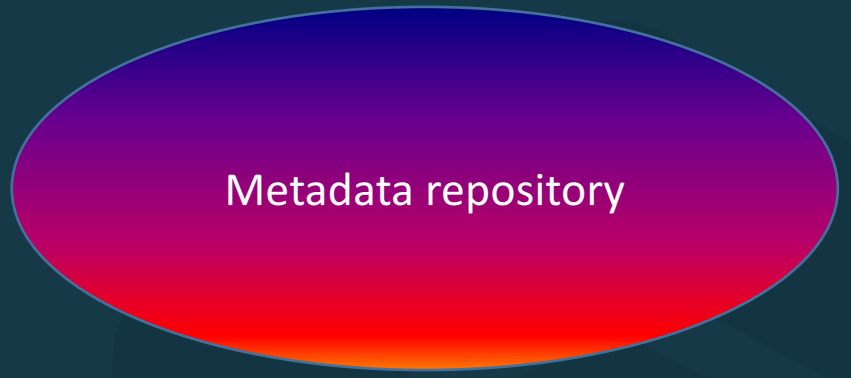
S-CUBED

An MDR controls definitions and versions of standards



S-CUBED

How does an MDR work?



Define (meta) data and definitions once – use again and again.

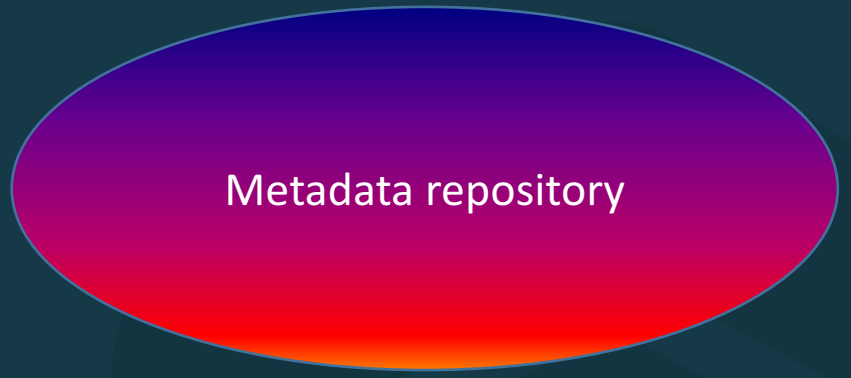
Control and understand changing versions of standards.

Be able to have user access and versioning of the metadata.

Be able to make impact analysis when standards, mappings or definitions change.

S-CUBED

Structure of Metadata



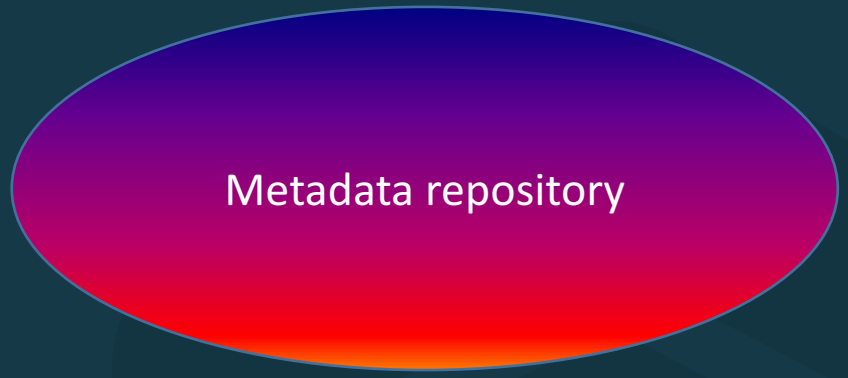
Tabular (Excel-like format)?

ODM?

Advanced graph data format ?

S-CUBED

Structure of Metadata



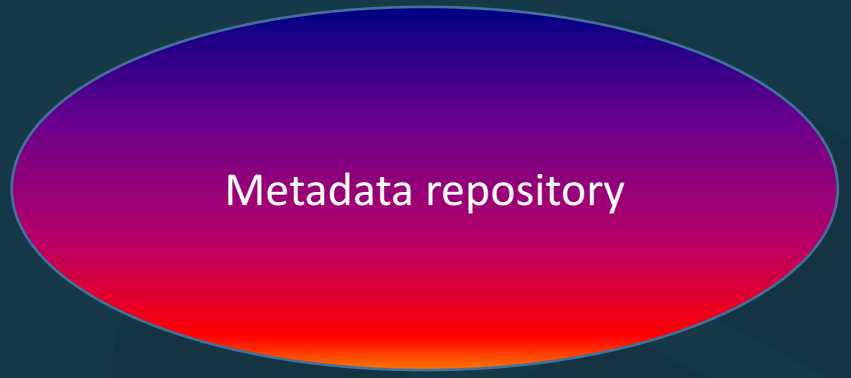
Optimal format is probably the Advanced graph data format…….

But we are much more familiar with tabular (Excel) format or maybe ODM in our industry.

Excel is too tabular. ODM is better – but is designed to hold CRF information (not metadata).

S-CUBED

MDR Software Solutions



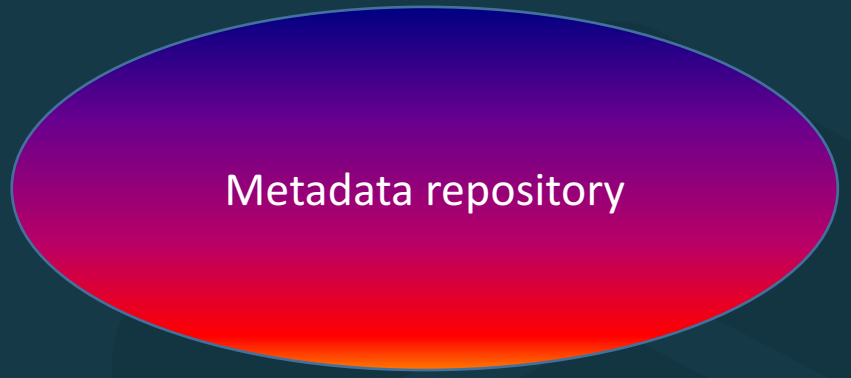
Many software products do **a part** of the job very well.

In my opinion the commercial FULL MDR solutions are not yet fully matured.

The industry has not yet developed and fully understood which metadata structure for clinical data is the optimal one.

S-CUBED

MDR Software Solutions

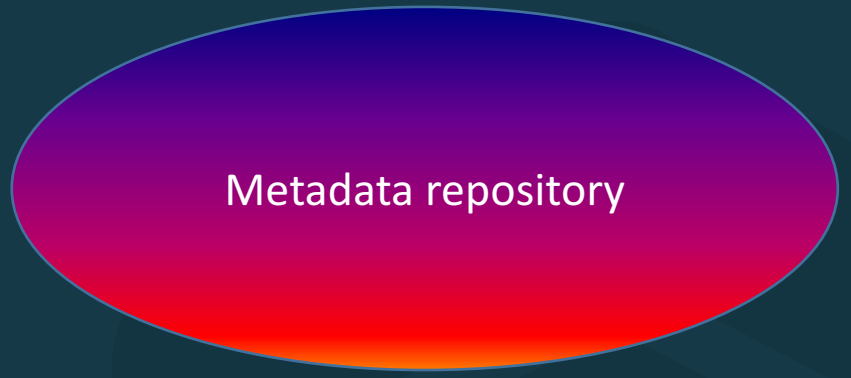


S-CUBED is co-developing the MDR solution GLANDON

(see more on https://www.a3informatics.com/)

S-CUBED

MDR Software Solutions



It is naïve to think you can make a “big-bang” IT project and tool implementation, where everything is understood and defined up-front and will just work.

Use your current XLS or ODM solutions and gradually expand and develop into a full solution. Learn as you go along.

MS Excel is good in many ways to experiment with. But it is usually not scalable – and it is not compatible for using the optimal metadata structure.

S-CUBED


IT'S BETTER TO SHOOT FOR THE STARS
AND MISS, THAN TO AIM FOR A PILE
OF MANURE AND HIT IT.

S-CUBED