



A GUIDE TO **CLINICAL DATA MANAGEMENT** FOR BIOTECH COMPANIES

CLINICAL DATA MANAGEMENT
ENSURING CLINICAL DATA READY
FOR SUBMISSION

iddi.com

CODE: SCDM 2024

THE NEED FOR CLINICAL DATA MANAGEMENT

CLINICAL DATA MANAGEMENT (CDM) IS THE PROCESS OF COLLECTING, CLEANING, AND MANAGING SUBJECT DATA IN COMPLIANCE WITH REGULATORY STANDARDS.



CDM AIMS TO GENERATE HIGH-QUALITY, RELIABLE AND STATISTICALLY SOUND CLINICAL DATA, COMPLIANT WITH PROTOCOLS AND REGULATIONS. SUCH DATA WILL BE USED FOR PROVING OR DISPROVING THE CLINICAL TRIAL HYPOTHESIS.

When properly handled, CDM significantly reduces time required for a new medical product launch

CDM processes ensure data integrity as defined by the FDA

01

Data integrity refers to the completeness, consistency and accuracy of data. Complete, consistent and accurate data should be Attributable, Legible, Contemporaneously recorded, Original or a true copy and Accurate (ALCOA)*

02

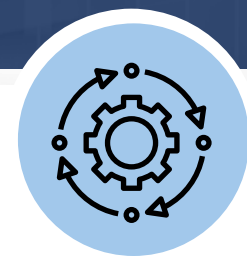
As data integrity has become an important topic in our industry, ALCOA+ has been cited as the ultimate reference. It adds the concepts that, in addition to ALCOA, data also needs to be Complete, Consistent, Enduring and Available.

* Data Integrity and Compliance with GCP – Guidance for the Industry – April 2016

THE DATA MANAGEMENT PROCESS

CDM activities start early in the clinical trial process, once the trial protocol, describing the study objectives and methodology, is designed.

THE CDM PROCESS BEGINS WITH THE END IN MIND, MEANING THE DATA MANAGEMENT ACTIVITIES ARE DESIGNED TO ANSWER THE RESEARCH QUESTION



TRIAL DESIGN

STUDY START-UP

STUDY CONDUCT

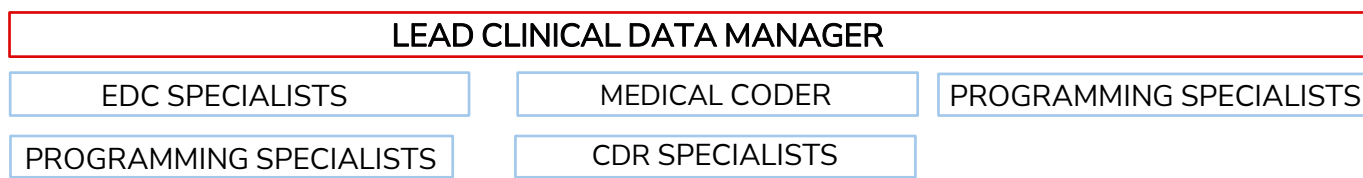
STUDY CLOSE OUT

STATISTICAL ANALYSIS

- Protocol Review
- Identification of data to be collected
- Data Management Plan (DMP)
- eCRF Design/EDC Set-up
- Database Design
- Training

- Data Capture
- Query Management
- Data Cleaning
- Data Reconciliation
- Medical Coding
- SAE Reconciliation
- Reporting

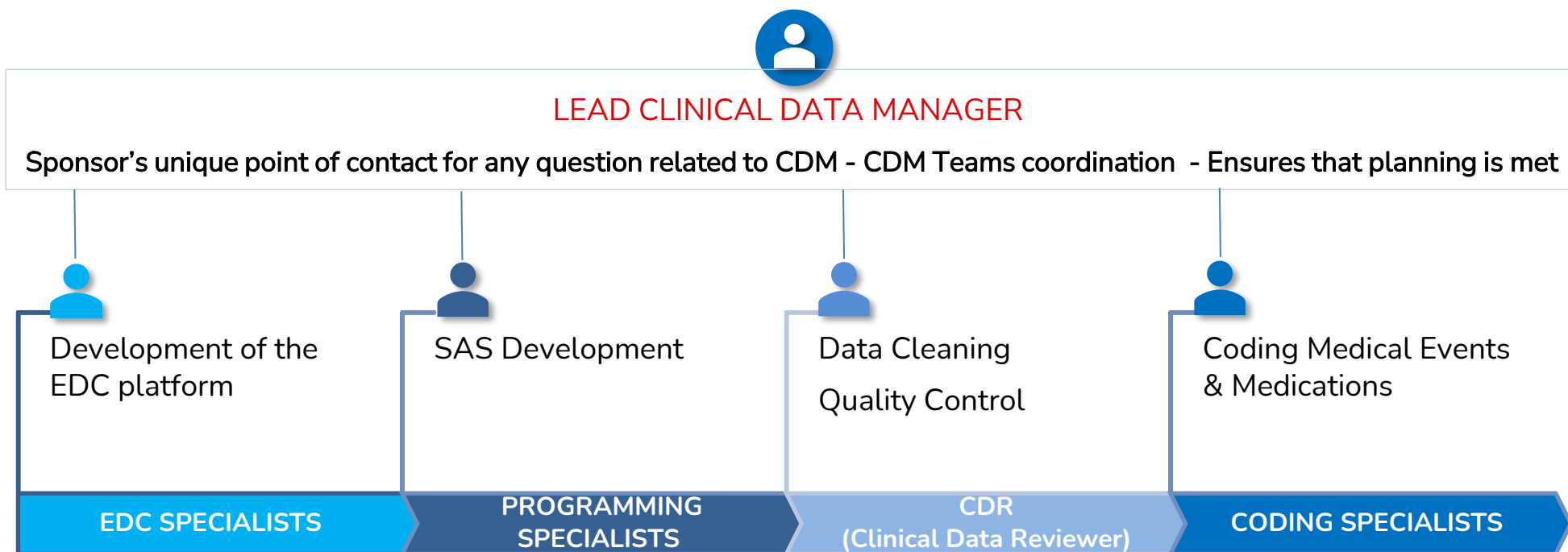
- Data Quality Review
- SAS Datasets for Analysis
- Database lock



THE CLINICAL DATA MANAGEMENT TEAM AT IDDI

TEAM MEMBERS OF CDM ARE ACTIVELY INVOLVED IN ALL STEPS OF A CLINICAL TRIAL FROM INCEPTION TO COMPLETION.

Clinical data managers should have adequate process knowledge in order to maintain the quality standards of the CDM processes. These are the reasons why the CDM team consists of different sub-teams based on their domain of expertise:



Study START-UP



EDC SYSTEMS ENSURE DATA INTEGRITY AND COMPLIANCE WITH REGULATORY STANDARDS

1. PROTOCOL REVIEW

The protocol is reviewed by CDM personnel with the aim of identifying the data items to be collected and the data items to be collected with respect to visits schedule.

These elements are entered into a Data Management Plan (DMP).

2. eCRF DESIGN

A Case Report Form (CRF) is designed by the CDM Team. Until early 2000, this was a manual process and data were collected on paper CRF. With the evolution of technology, Electronic Data Capture (EDC) systems have been developed.

EDC systems include:

- The electronic Case Report Forms (eCRF) : a digital questionnaire that is used to collect data about a clinical study and research participants.
- The electronic Data Validation Plan (eDVP)
- Online listing and reports.

The EDC set-up activities consist of

- The creation of the different forms of the eCRF in accordance with the study protocol which will allow data entry by site staff
- The programming of the data collection flow and edit checks
- The development of the necessary online listings and metrics reports

3. VALIDATION

Validation proves that the EDC system is configured correctly. Validation is a procedure for testing and checking all system functions. Its purpose is to confirm that the EDC is configured correctly and meets all requirements.

4. TRAINING

Once the EDC system setup is final, training will be provided to site users and monitors to ensure data will be entered appropriately in the EDC system.

Study CONDUCT

DURING THIS PHASE INVESTIGATORS AND SITE PERSONNEL WILL CAPTURE DATA USING THE EDC SYSTEM.



CDM Team manages this conduct phase by:

1. DATA REVIEW & CLEANING

Regularly reviewing and cleaning the collected data to ensure data consistency and accuracy

2. REPORTING

Reporting to the Sponsor on the progress of the study

3. CODING

Coding the medical events and medications

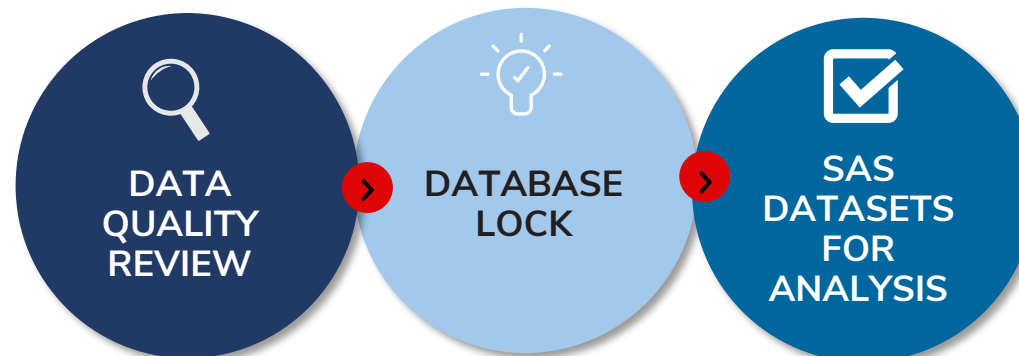
4. INTEGRATION IN THE EDC SYSTEM

Integrating the external data (such as randomization data, Labs data, eCOA, etc.) into the EDC system

5. LISTINGS & REPORTING

Providing data listings & reports to Sponsor for medical review or to Data Safety Monitoring Board

Study CLOSE OUT



At the end of the trial,

- The database is locked (all user accesses are closed to avoid further data collection)
- Data are exported and mapped into the appropriate SAS structure and transferred to the Biostatistics team for analysis. Mapping is the process of transforming the structure of the EDC database into a SAS structure that will be used by Biostatisticians, usually the CDISC SDTM structure.

REGULATIONS, GUIDELINES & STANDARDS IN CDM

THIS PARAGRAPH LISTS REGULATIONS, GUIDELINES AND STANDARDS THAT ARE CONSIDERED THE MOST RELEVANT REFERENCES THAT MAY BE USED DURING A CLINICAL TRIAL. IT IS A NON-EXHAUSTIVE LIST.

FDA

Code of Federal Regulations (CFR), [21 CFR Part 11](#): This regulation applies to records in electronic format that are created, modified, maintained, archived, retrieved, and/or transmitted.

ICH

- [E6\(R2\)](#): Integrated Addendum to ICH
- [E6\(R1\)](#): Guideline for Good Clinical Practice
- [E8](#): General Considerations for Clinical Trials
- [E9](#): Statistical Principles for Clinical Trials

WORLD MEDICAL ASSOCIATION

[Declaration of Helsinki](#): Ethical principles for medical research Involving Human Subjects – Version Oct 2013

CLINICAL DATA INTERCHANGE CONSORTIUM (CDISC)

CDISC is a multidisciplinary non-profit organization that has developed standards to support acquisition, exchange, submission and archival of clinical research data and metadata (metadata is the data about the data collected).

Among the standards, two important ones are the [Study Data Tabulation Model \(SDTM\)](#) and the [Clinical Data Acquisition Standards Harmonization \(CDASH\)](#).

- **The SDTM standard** describes the details of model and standard terminologies for the data and serves as a guide to the organization.
- **CDASH** defines the basic standards for the collection of data in a clinical trial and enlists the basic data information needed from a clinical, regulatory and scientific perspective.

SOCIETY FOR CLINICAL DATA MANAGEMENT

Good Clinical Data Management Practices ([GCDMP](#))

THE ELECTRONIC DATA CAPTURE TOOL

ELECTRONIC DATA CAPTURE (EDC) SYSTEMS ARE WEB-BASED SOFTWARE APPLICATIONS USED TO **COLLECT, CLEAN, TRANSFER AND PROCESS** DATA IN CLINICAL TRIALS.



IDDI uses state-of-the-art systems such as ID-base (EvidentIQ Marvin) & Medidata RAVE

 ID-base  Rave

The use of EDC systems has become essential, especially in multi-centric trials, to handle the huge amount of data collected while maintaining quality data.

- In regulatory submission studies, maintaining an audit trail* of data management is of paramount importance.
- Many software tools are available for data management.
 - In terms of functionalities, most software tools are rather similar.
 - The differences are generally found in the Graphic User Interface, the integration with other systems such as RTSM systems and the reporting tools.
 - The ultimate difference is the DM team who has the necessary scientific background and experience to help you set this EDC up!

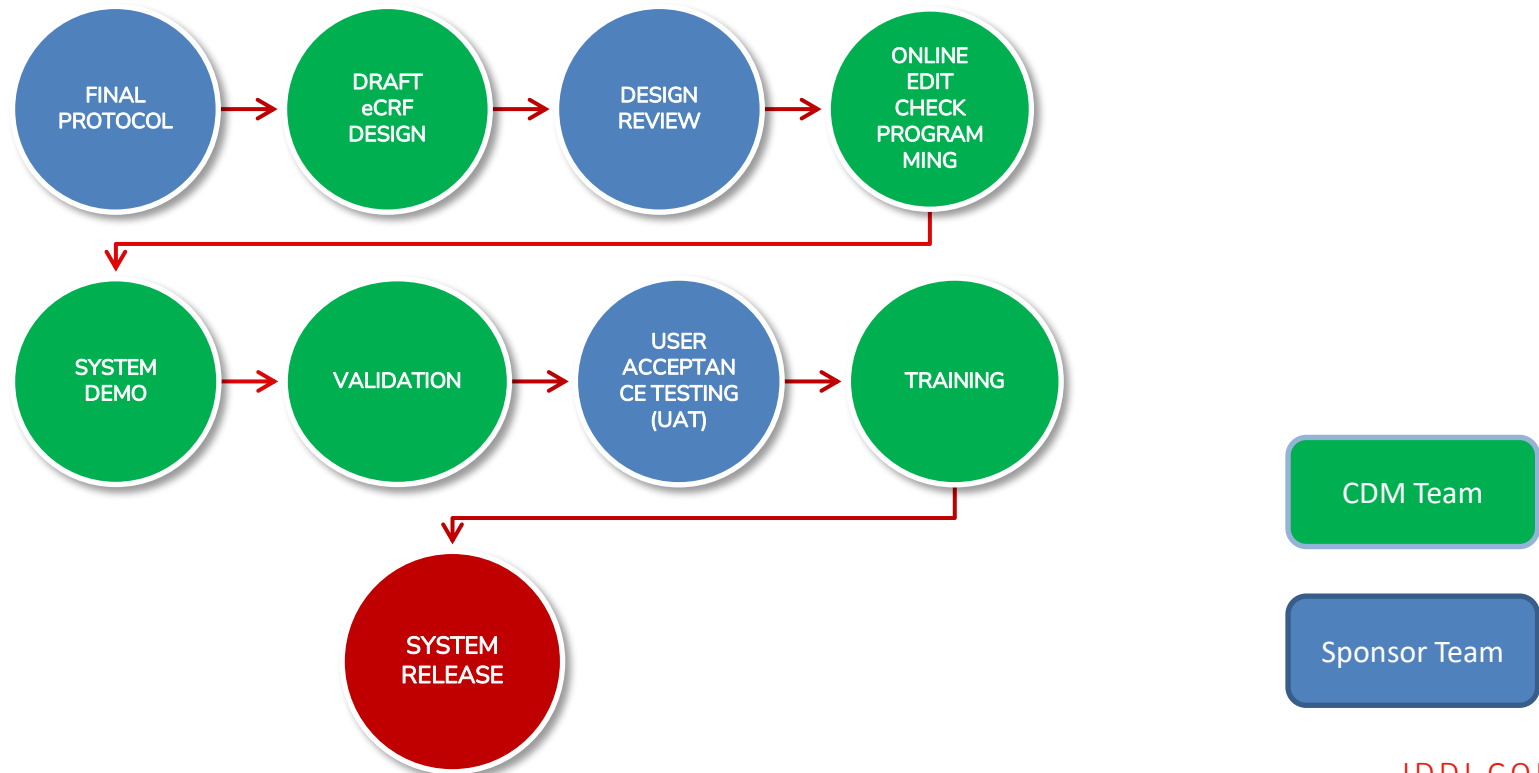
“ **EDC SYSTEMS ENSURE DATA INTEGRITY AND COMPLIANCE WITH REGULATORY STANDARDS** ”

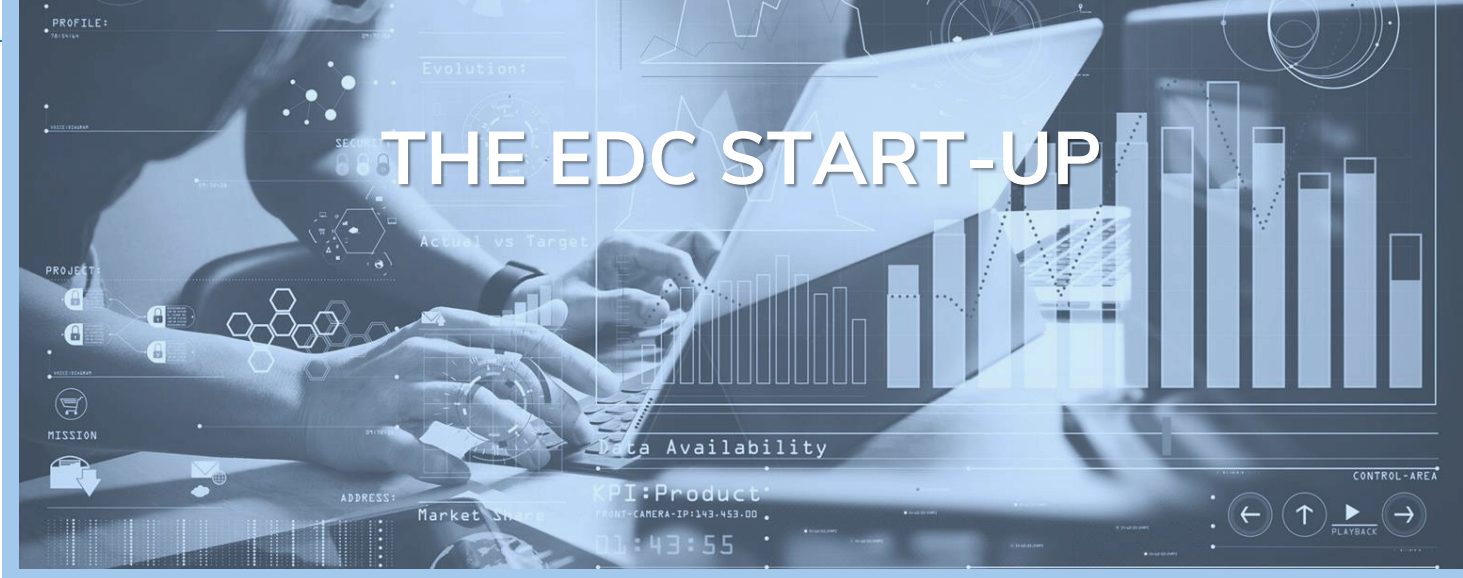
* FDA definition: A secure computer generated, time stamped electronic record that allows reconstruction of the course of events relating to the creation, modification, and deletion of an electronic record

THE EDC START-UP

ONCE THE PROTOCOL HAS BEEN FINALIZED, THE ELECTRONIC DATA CAPTURE SYSTEM CAN BE SET-UP AND CUSTOMIZED TO COLLECT, MAINTAIN AND ARCHIVE DATA IN ACCORDANCE WITH THE PROTOCOL AND APPLICABLE STANDARDS.

The EDC setup process requires close collaboration between the CDM Team and the Sponsor:





1. DESIGN OF THE FIRST DRAFT OF THE eCRF

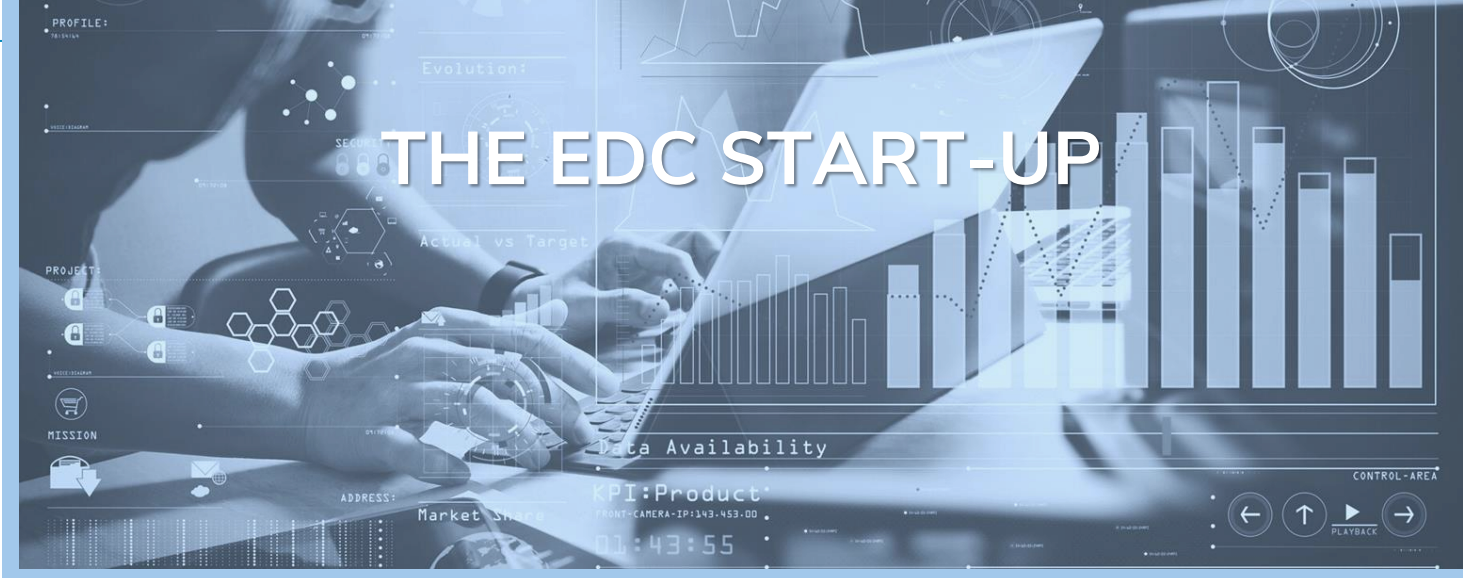
The eCRF consists of a sequence of visits in which the different assessments - as defined in the study protocol - are designed. This step ensures all the data needed for the statistical analysis are collected but also, prevents data duplication and facilitates the conversion to SAS (Statistical Analysis Software), datasets that will be used by the Biostatisticians. The first draft CRF will be used as basis for the review with the Sponsor

2. DESIGN REVIEW

During the design review Sponsor, Biostatistician and CDM Team will ensure the design of the different visits / assessments is in accordance with the trial needs and that data collected will allow to test the trial hypotheses and answer the research question.

3. PROGRAMING

The data fields should be clearly defined and be consistent. For that purpose, programming of the edit checks will prevent as much as possible, the entry of data in wrong forms or the entry of inconsistent data – this helps decrease the time needed for data cleaning and improves data quality.



4. FINAL DESIGN REVIEW

Once a final draft of the eCRF is available, the Sponsor will review it a last time before validation and acceptance.

5. VALIDATION

For computer systems used in clinical trials, computer system validation (CSV) is a key regulatory requirement. CSV is performed to demonstrate that the whole system, including software, hardware process, and people, performs consistently and is fit for purpose. A validation plan is developed, that will outline the testing required to meet the requirements. This essential task is performed by a team independent of the set-up team.

6. USER ACCEPTANCE TESTING (UAT):

Following system validation, the EDC is made available to the Sponsor to perform User Acceptance Testing. This testing should confirm that the system is working according to the specifications.

7. TRAINING

The Lead CDM provides training of the EDC to the Investigators/monitors - ideally face to face.

8. SYSTEM RELEASE

Release of the system to enable the sites to start data entry. The EDC release is done few days before the First Patient First Visit.

REPORT AND LISTINGS

Reports and listings are an important component of the EDC platform as it will allow the Sponsor to closely monitor the progress of the data collection activities and data entries by sites.



Reports and listings are agreed upon prior their release and are made available online. These reports and listings also go through system validation.

01

Reports are metrics used to follow-up on the progress of the study and the performance of the sites.

02

A listing is a simple dump of the data of one form of the eCRF to enable the review of the specific data of that form (e.g Demographic data)

REPORT AND LISTINGS

EXAMPLES

Query Status per patient							
Site ID	Subject study ID	Study Phase	Cohort	Is patient enrolled in Phase 1 dose expansion?	# Open queries	# Answered queries	# Closed queries
001	001-9001	Phase 1	20 mg/kg (Q2W)	No	32	0	328
001	001-9002	Phase 1	20 mg/kg (Q2W)	No	22	2	232
001	001-9003	Phase 1	20 mg/kg (Q2W)	No	27	5	246
001	001-9006	Phase 1	30 mg/kg (Q2W)	No	14	0	85
001	001-9009	Phase 1	50 mg/kg (Q4W)	No	8	0	122
001	001-9011	Phase 1	50 mg/kg (Q4W)	No	39	0	135
001	001-9012	Phase 1	50 mg/kg (Q4W)	No	33	0	151
001	001-9018	Phase 1	50 mg/kg (Q4W)	Yes	12	0	235
001	001-9022	Phase 1	50 mg/kg (Q4W)	No	6	1	93
001	001-9031	Phase 1	50 mg/kg (Q4W)	Yes	13	0	212
001	001-9034	Phase 1	50 mg/kg (Q4W)	Yes	33	0	258
001	001-9036	Phase 1	50 mg/kg (Q4W)	No	27	1	153
001	001-9041	Phase 1	50 mg/kg (Q4W)		1	0	0
001	001-9043	Phase 1			3	0	0
001	001-9046	Phase 1			2	0	0
001	001-9047	Phase 1	40 mg/kg (Q4W)	Yes	49	0	149
001	001-9048	Phase 1			1	0	0
001	001-9049	Phase 1	40 mg/kg (Q4W)	Yes	31	1	184
001	001-9050	Phase 1	40 mg/kg (Q4W)	Yes	46	5	143
001	001-9057	Phase 1			4	0	0
001	001-9061	Phase 1			3	0	0
001	001-9065	Phase 1			2	0	0
001	001-9087	Phase 1		Yes	2	0	0

Time to Data Entry Report										
Site ID	# Total of Recruited Patients	# Total of Treated Patients	# Total of Events taken into account	Mean Days	Less than 7 Days	7 - 13 Days	14 - 20 Days	21 - 28 Days	29 Days - 3 Months	More than 3 Months
001	23	23	264	33	163	38	9	9	17	28
003	22	22	144	13	79	36	12	6	6	5
004	14	14	123	8	98	14	7	0	0	4
005	2	2	2	144	1	0	0	0	0	1
006	3	3	26	63	6	1	3	2	5	9
007	6	6	112	26	45	14	12	12	19	10
009	3	3	32	8	16	11	3	1	1	0
011	8	8	132	18	74	30	7	4	9	8
305	1	1	14	19	3	1	2	5	3	0
306	3	3	26	5	20	4	0	1	1	0
607	2	2	2	0	2	0	0	0	0	0
610	2	2	6	0	6	0	0	0	0	0
611	1	1	2	7	1	0	1	0	0	0
TOTAL	90	90	885	26	514	149	56	40	61	65

DOCUMENTATION

During the system set-up, the following documents are developed by the CDM team:

- **Data Management Plan - DMP*** - This document describes the different steps of the CDM process and roles throughout the lifecycle of the clinical trial
- **EDC specifications:** This document describes the configuration of the EDC system
- **eCRF:** this is an electronic document designed to record all of the information which is required by the Protocol. The CRF is the paper version of the eCRF
- **eDVP** is a document that contains a detailed description of all the actions programmed, of the edit checks and the calculations for derived variables
- **CRF completion guidelines.** This document describes how to enter data into the system. It is intended to assist the investigator to complete the eCRF in systematic manner
- **EDC reports specifications** provide a mock example of each report as well as a description of the forms and fields that will be used to develop the report, and the conditions that will be applied

SAS DATABASE & CLEANING TOOLS SET-UP

SEVERAL ACTIVITIES CANNOT BE PERFORMED DIRECTLY WITHIN THE EDC. THEREFORE, PROGRAMMING, USING SAS SOFTWARE IS REQUIRED:

FOR STRUCTURING THE DATA COLLECTED IN THE EDC TOOL:

- The **Clinical Data Interchange Standards Consortium (CDISC)** has created standards that are now mandatory for regulatory submission.
- **Study Data Tabulation Model (SDTM)** is one of these standards. Data extracted from the EDC need to meet this CDISC SDTM Standards.
- SDTM provides a uniform standard from study to study to ease data exchange ly and across vendors.
- It is highly recommended to use this structure for all trials to be submitted to Regulatory Authorities

FOR COMPLEMENTING LISTINGS, REPORTS, CHECKS DEVELOPED IN EDC.

- Complex edit checks, Patients profiles and complex reports will be developed in SAS.
- The CDM team needs to develop a SAS Data Validation Plan DVP to complement the eDVP.
- The SAS DVP also defines data not collected directly in the EDC platform such as laboratory data, imaging data, MedDRA codes and/or WhoDrug codes

FOR RECONCILING EXTERNAL DATA

- Data such as laboratory data, imaging data, pharmacovigilance data, etc. are collected in different systems and need reconciliation.
- It is important to make sure the identifiers of such data correspond to the identifiers in the EDC to allow correct integration of those data into the SAS mapped database.

FOR CODING

- For importing the terms to be coded into the coding system and
- For exporting the codes from the coding system into the SAS database



TIMING OF AVAILABILITIES OF SAS DELIVERABLES

- **MAPPING PROGRAM IS READY 3 MONTHS AFTER EDC SYSTEM RELEASE**
- SAS cleaning tools are delivered on a date agreed with the Sponsor based on the trial needs

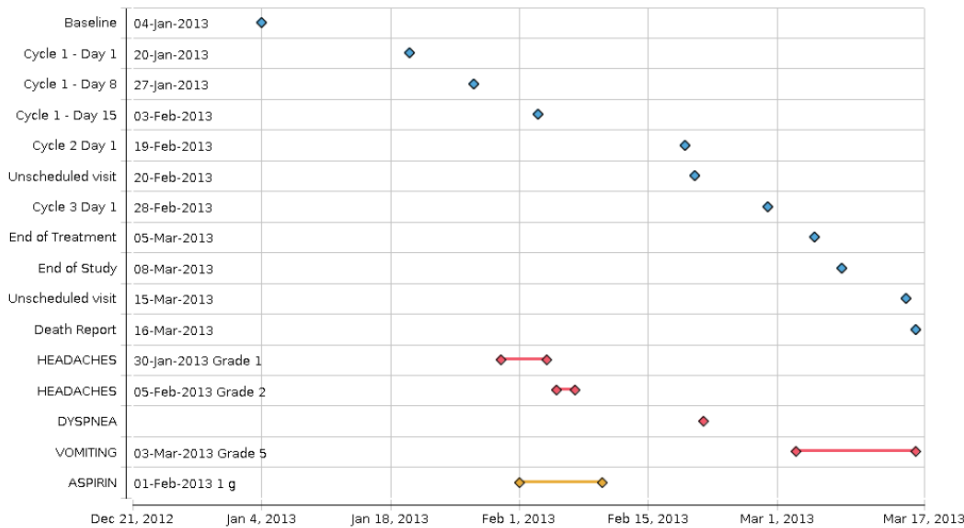
PATIENT PROFILES

EXAMPLES



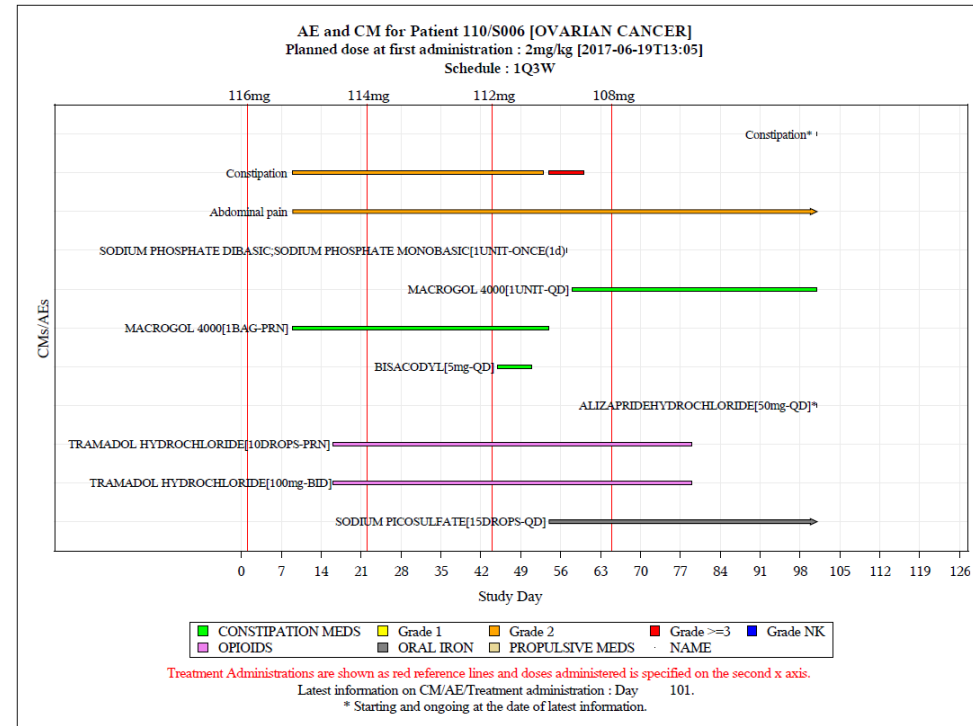
Results: Patient Profile Report - DRAFT

Patient: 99-99-02 Initial Dose: 80 mg Daily



	Jan 30, 2013	Feb 1, 2013	Feb 5, 2013	Feb 21, 2013	Mar 3, 2013
2_AE	DYSYPNEA			21-Feb-2013 Grade 2	
	HEADACHES	30-Jan-2013 Grade 1	05-Feb-2013 Grade 2		
	VOMITING				03-Mar-2013 Grade 5
3_CM	ASPIRIN	01-Feb-2013 1 g			

May 12, 2015, 8:45 AM



SAS DATABASE & CLEANING TOOLS SET-UP

The following additional documents are developed during this SAS database set-up activity :

- Annotated CRF*: makes the link between the field from the CRF and the variable in the SAS database
- SAS DVP*
- External data transfer specifications
- Serious Adverse Events (SAE) reconciliation plan*

**Annotated CRF: An annotated CRF is generally defined as a blank CRF with markings, or annotations, that coordinate each datapoint in the form with its corresponding dataset name. Essentially, an annotated CRF communicates where the data collected for each Question is stored in the database.*

**SAS DVP*: list of all checks that will be programmed in SAS or performed manually based on SAS listings.*

**SAE Reconciliation*: The process of comparing key safety data variables between the drug or device safety SAE database and the clinical database in order to identify any discrepancy, determine whether a discrepancy is acceptable or not and, if acceptable, document the discrepancy.*

SDTM ACTIVITIES

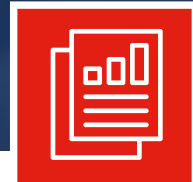


CDISC and FDA have worked together closely since CDISC's inception to ensure data standards allow regulatory reviewers to receive, process, review and archive submissions more effectively.

- These standards provide a consistent general framework for organizing study data, including templates for datasets, standard names for variables and standard ways of doing calculations with common variables.
- **CDISC standards are required for regulatory submissions to FDA.** You can read the following on the FDA website :”*FDA may refuse to file for New Drug Applications (NDAs) and Biologics License Applications (BLAs) any electronic submission whose study data do not conform to the required standards*”– EMA will follow soon the same standards. Therefore, no need to say that **having a SAS database 100% compliant with SDTM is a key of success in case of file submission to Authorities.**

Even if your CRF is designed to already comply with CDISC requirements (CDASH), even if most of the EDC vendors claim that the data extracted from their EDC system follow SDTM, transforming the EDC data structure into SDTM is a step that can't be skipped.

SDTM ACTIVITIES



The SDTM mapping process is usually as follows:



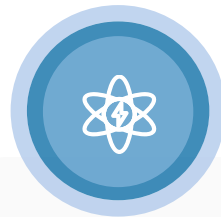
ANNOTATION OF THE CASE REPORT FORM

- Identify on the CRF the data to be mapped
- Identify the SDTM datasets to which each fields of the CRF will belong
- Identify custom domains for data without matching SDTM datasets



DEVELOPMENT OF THE SAS MAPPING PROGRAM BASED ON THE ANNOTATED CRF

- Structure (including EPOCH, Baseline Flag,....)
- Trial Design Datasets, Supplemental Qualifier Datasets and Finding About dataset
- Terminology



QC OF THE SAS PROGRAMMING

QC is done by a second independent programmer who develops also the program based on the same annotated CRF, and then compare the outputs until no more discrepancies



COMPLIANCE

Ensuring compliance with the CDISC SDTM requirements by using the Pinnacle 21 tool

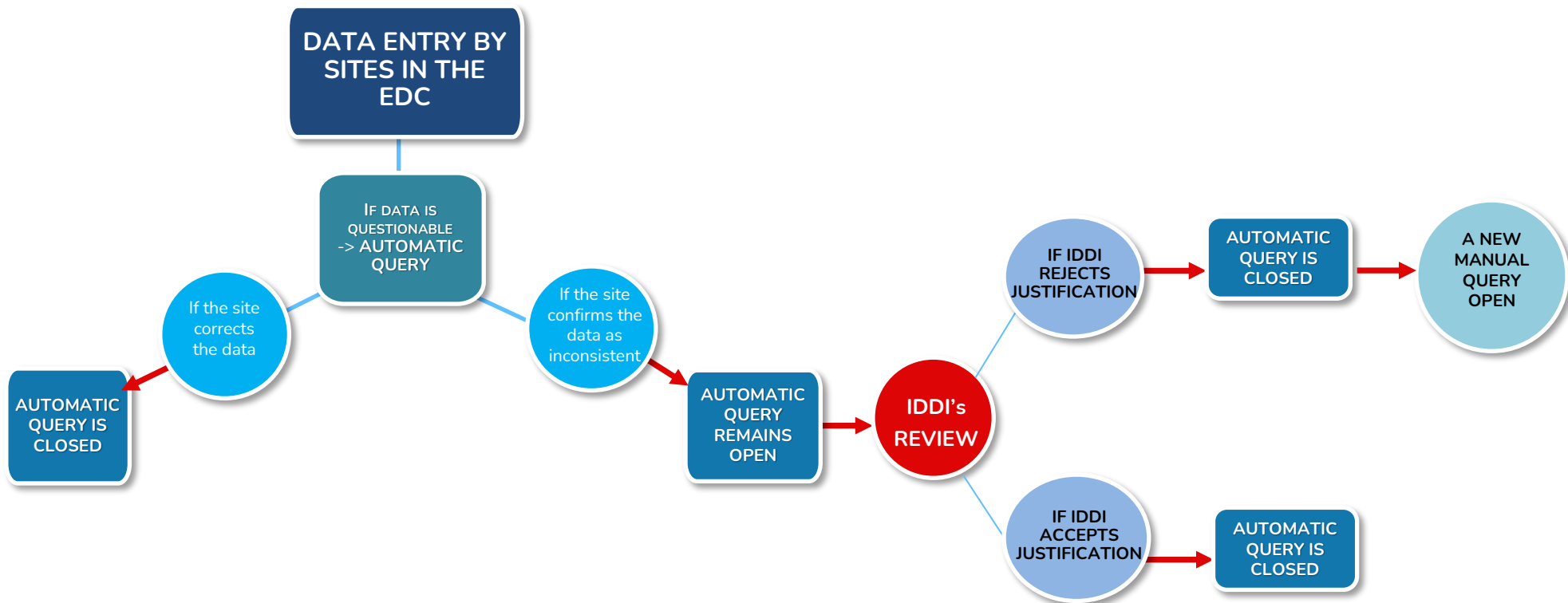


DOCUMENTATION

Development of the documentation needed for the regulatory submissions (Data Reviewer's Guide, Define XML, Define PDF)

CLEANING ACTIVITIES

PROCESS



CLEANING ACTIVITIES

PROCESS



THE AUTOMATIC CLEANING ACTIVITIES START ONCE SITES START ENTERING DATA IN THE EDC.

- When the sites enter data, the checks develop in EDC run automatically. In case data are wrong or inconsistent a question is automatically posted in the EDC (automatic query).
- If the site changes the data to make it correct, the query is automatically closed. But if the site confirms the data as inconsistent, the query remains.
- IDDI reviews the site's justification and either accepts the justification and closes the query or doesn't accept it and then closes the automatic query but raises a new query more detailed (manual query).

THE MANUAL CLEANING STARTS ONCE THE NEEDED SAS PROGRAMS ARE READY.

- It is split between the Sponsor/Medical reviewer and IDDI.
- IDDI does all the cleaning that does not require medical expertise. The split of the cleaning activities is agreed with the Sponsor as well as the frequency of this cleaning.
- IDDI Manual cleaning is based on the SAS DVP. Queries are posted manually in the EDC system and followed-up until resolution

AUTOMATIC AND MANUAL CLEANINGS ARE DONE UNTIL THE DATABASE LOCK.

- To follow-up on the cleaning, the main tool is the 'Clean Patient tracker'.
- It shows the status of the patient regarding all the cleaning activities (IDDI, Sponsor, Medical reviewer).

CODING ACTIVITIES

MEDICAL CODING HELPS IN IDENTIFYING AND PROPERLY CLASSIFYING MEDICAL TERMS USED IN THE ECRF INTO INDUSTRY STANDARD DICTIONARY TERMS.



Medical Coding aims to achieve data consistency and avoid unnecessary duplication. This will ensure that data can be efficiently analyzed by Biostatisticians.



Medical Dictionary for Regulatory Activities (MedDRA) is used for the coding of adverse events as well as other illnesses



World Health Organization – Drug Dictionary Enhanced (WHO-DDE) is used for coding the medications

During the trial CDM Team codes the data and Sponsor reviews it on an agreed frequency. At the end of the trial Sponsor will formally approve all the coding performed by CDM Team.

MedDRA example:

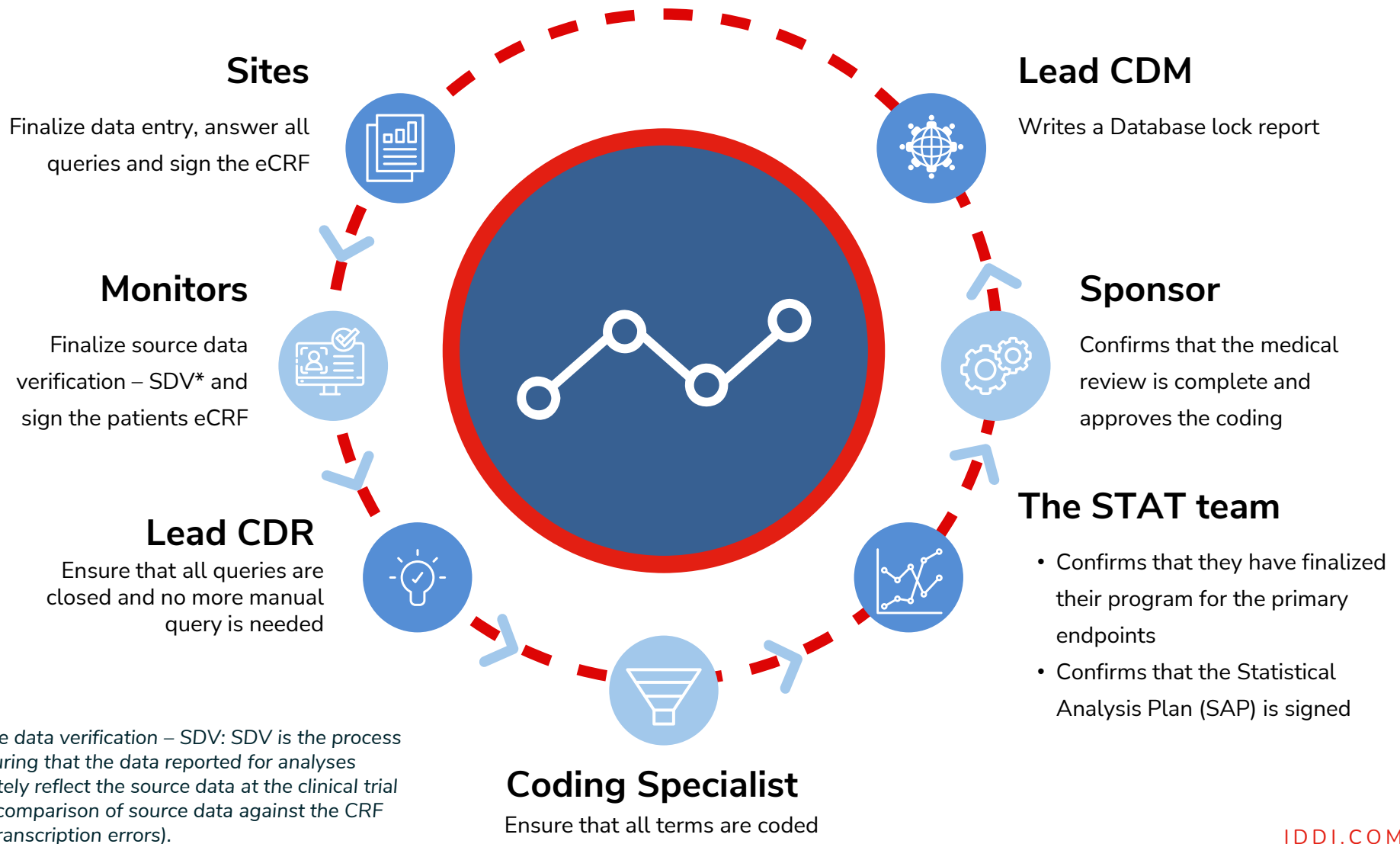
- Terms reported in eCRF: 'LOW BACK PAIN', 'LOWER BACK PAIN', 'LOWER BACK PAIN RIGHT SIDED', 'RIGHT SIDE LOW BACK PAIN'
- > Can all be coded using MedDRA LLT 'Low back pain'

WHO-DDE example:

- Medications reported in eCRF: 'ASPIRINE', 'ASPIRIN', 'ACETYLSALYCILIC ACID'
- > Can all be coded using WHO drug name 'Acetylsalicylic Acid'

DATABASE LOCK Activities

Once Last Patient Last visit (LPLV) is reached, CDM will ensure that the database is ready to be locked.



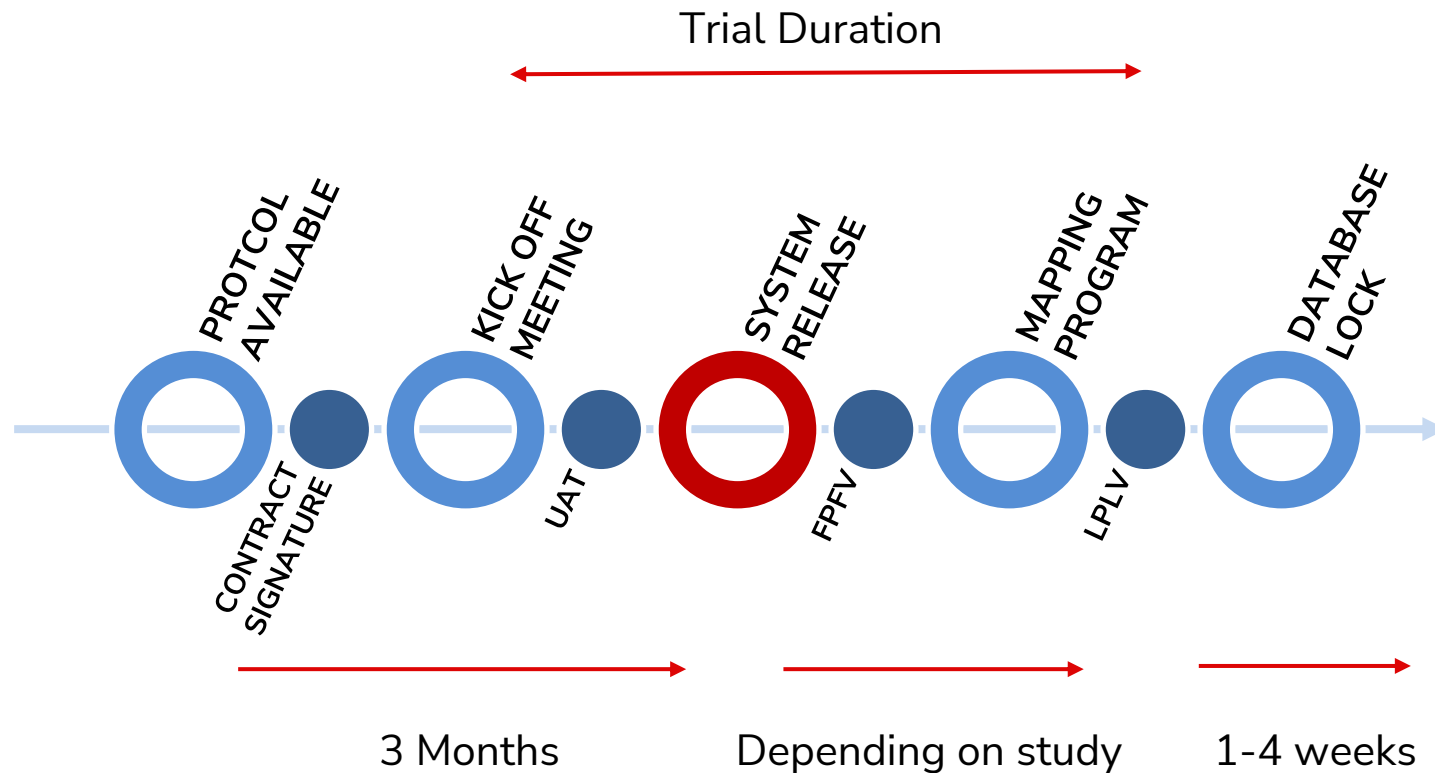
Database Lock ACTIVITIES and Study End



THE LOCK OF THE EDC SYSTEM CONSISTS OF REMOVING THE WRITE ACCESS TO THE ECRF – ALL USERS WILL HAVE READ ONLY PERMISSION.

- When the EDC specialist has changed the permission, the Lead CDM generates the final version of the SAS datasets.
- The locked datasets are provided to Biostatisticians for analysis.
- Once the final statistics results are available, CDM will provide the CRF data in PDF format to the sites (directly or via the Sponsor).
- Once confirmation of receipt is received, the trial will be removed from the EDC system.
- IDDI receives from the EDC vendor all the files required to be able to restore the trial in the EDC if needed.

TIMELINES



BIostatistics & Data Management: Combined Services Ensuring You Meet Your Study Goals

CONCLUSION



BIostatistics & Data Management: Combined Services – Advantages



ENSURES THAT ALL IMPORTANT ENDPOINTS ARE INCLUDED FOR YOUR STUDY.

ENSURES DATA IS COLLECTED IN THE BEST WAY TO MEET THE GOALS OF YOUR STUDY.

STREAMLINES THE CREATION OF THE FINAL DATASETS FOR ANALYSIS AND SUBMISSION.

AVOIDS DELAYS IN REPORTING ANY DATA ISSUES TO YOUR CLINICAL DATA MANAGEMENT TEAM.

ENSURES THE QUALITY OF THE DATA USED IN YOUR STATISTICAL ANALYSES.

ENSURES YOUR DATABASE IS LOCKED ON TIME & BE ABLE TO SUBMIT RESULTS SHORTLY AFTER THE DATABASE LOCK.



PASSION. SCIENCE. EXPERIENCE.

IDDI

INTERNATIONAL DRUG DEVELOPMENT INSTITUTE



IDDI.COM



info@iddi.com

Biostatistical & eClinical Services since 1991