PharmaSUG 2023 - Paper SI-006

Digital Data Flow (DDF) and Technological Solution Providers

Prasoon Sangwan, Piyush Singh
TATA Consultancy Services Ltd.

ABSTRACT

Digital Data Flow (DDF) is an initiative to organize and automate the processing of clinical data and study protocol. From a technology perspective, one of the key purposes of this initiative is to deliver the technical standards which can be utilized to mechanize the study execution process, create a flexible solution and minimize manual effort during the study life cycle. One of the most important principles of the DDF initiative is being vendor agnostic, which means that different organizations can implement their solution in their own way, using reference architecture (RA) from DDF, from both process and technology perspectives.

This paper explains how the technology providers/ technological product vendors can utilize the DDF deliverables to help pharmaceutical companies with new solutions/platforms to innovate and automate their manual and traditional study execution process which ultimately can help to reduce overall cost, duration of the study and operational effort, and increase the return. This paper also explains how Pharma companies can utilize the strength of technology to take maximum advantage of DDF.

INTRODUCTION

Digital Data Flow is an initiative for pharmaceutical and clinical research aimed at providing a Reference Implementation Architecture model supported by a Study Definition Repository that digitizes information in an industry-standard format. This initiative enables the creation of a digitized protocol that downstream systems involved in the execution of the clinical trial can consume. This implementation improves the quality and timeliness of clinical information flow while maintaining traceability across systems, leading to improved and optimized processes throughout the clinical trial life cycle.

The open-source, vendor-agnostic Reference Implementation architecture helps technology providers leverage the standardized data model to induce interoperability and automation into data exchange methodology. This can revolutionize information flow and enable analytics to optimize the process right from the protocol design. The multi-stakeholder governance model will help the solution sustain against the changing technological landscape.

CLINICAL TRIAL INFORMATION FLOW

The protocol is created using the Clinical Development Plan and identified Clinical Endpoints. The protocol serves as the primary source of information for the Statistical Analysis Plan (SAP), Clinical Trial Management System (CTMS), data collection systems, Study Reference Manuals, Clinical Study Report (CSR), Local Registrations and approvals, and Trial Master File (eTMF). Currently, the protocol is either created in a word document or filled in predefined templates, making it challenging to extract all the information and requiring significant human intervention.

The data from the protocol is transferred to other systems manually, leading to inconsistency, redundancy, and incomplete and delayed information. The lack of interoperability between systems requires each process to be executed manually and linearly, such as defining Schedule of Activities (SoA), SAP, setting up the EDC (Electronic Data Capture).

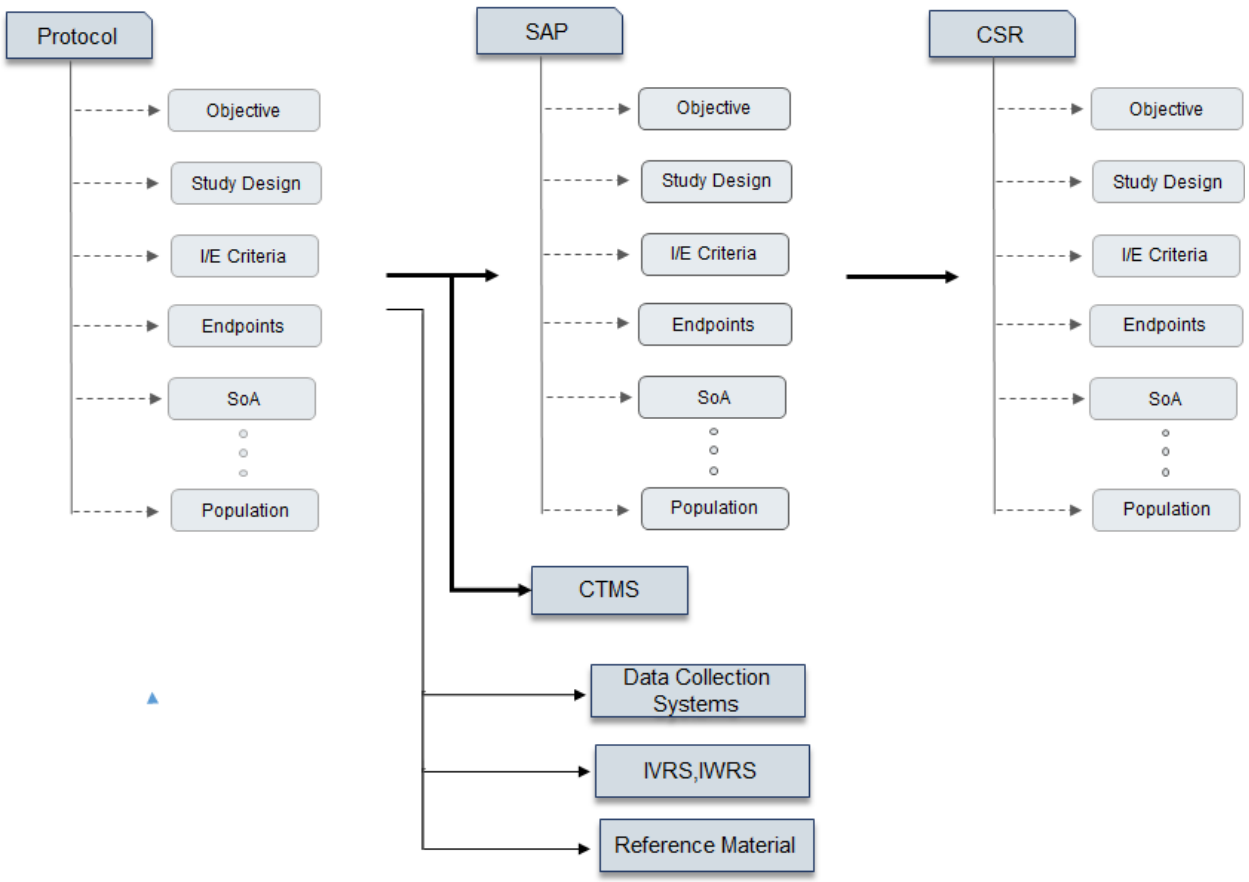

Figure 1. Clinical Information Flow

Any amendment to the protocol becomes a bigger challenge as there is no easy way to analyze the impact. The manual process leads to operational inefficiencies, as well as potential quality and compliance issues.

The overall concept of the DDF initiative is to create Study Design Metadata that can be used instead of the clinical protocol. Earlier, data standards were only coming into the picture during data storage, but the plan is to create metadata for the entire study life cycle. The Study Design Metadata will not only help with data collection but also throughout the study execution, process automation, and artificial intelligence.

Standards are very much required for most automated solutions because automation requires machine-readable and executable formats. The system can be built/coded to deal with the standard format so that custom changes can be avoided as much as possible. If we get well-defined standards, technology providers/ IT companies can create solid and stable platforms to create end-to-end systems based on the given data standards, and this can be used across the pharmaceutical industry. In the current era of cloud IT system can be designed in multiple modules like source, target, interim execution etc. If there is already a interface available for such system, target/downstream applications can interact using the application programming interface (APIs) to the current interface.

DDF envisions the data flowing automatically and parallels in an interoperable way between the systems in the form of Study Design Metadata, to maintain end to end traceability as described in Figure 2. This makes the information in all the systems consistent leading to efficient execution with quality and compliance.

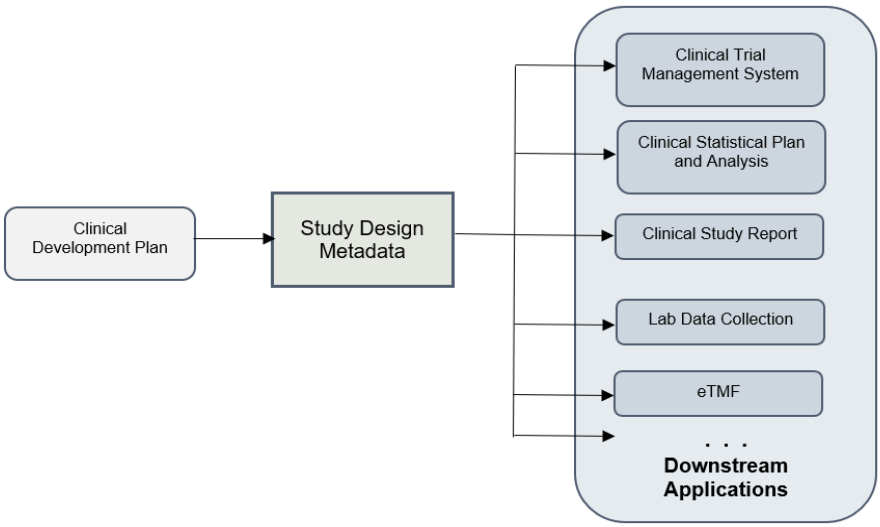

Figure 2. Clinical Data Flow with DDF

KEY COMPONENTS OF DDF

DDF comprises upstream and downstream systems that are interconnected through a Study Definition Repository (SDR), utilizing data and technology standards. Technology providers (maybe outsider IT companies or in-house IT divisions) play a critical role in DDF by ensuring the maintenance of data standards through proper governance models enforced via the underlying technology. It is crucial for them to implement vendor-agnostic and open-source solutions to facilitate rapid and effortless adoption by the pharmaceutical industry. This will make the solution scalable and flexible, enabling it to adapt to frequent changes while adhering to an agile approach. The key components utilized are as follows:

1. SOURCE/UPSTREAM SYSTEM

In the clinical digital data flow, the source system consists of a protocol authoring tool used to generate a digitized protocol. The source system utilizes the standards from the Study Definition Repository (SDR) to capture information related to major protocol components such as study information, study objectives, endpoints, inclusion/exclusion criteria, schedule of activities, and more.

2. STUDY DEFINITION REPOSITORY (SDR)

In the context of clinical digital data flow, the Study Definition Repository (SDR) is one of the key components of DDF. It serves as a middle layer between study definitions/protocols and study execution. The SDR utilizes technology and data standards to efficiently manage the flow of information between different tiers. Downstream applications can interact with the SDR through an application programming interface (API), which connects to the current interface.

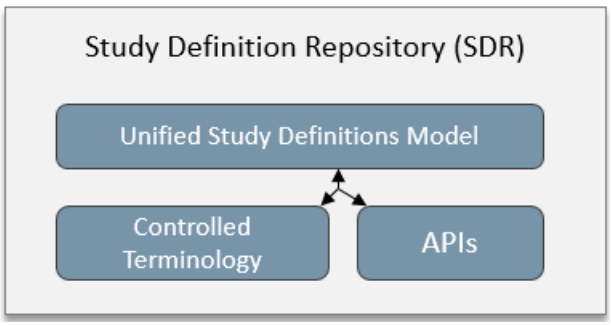

Figure 3. Study Definition Repository

2.1. UNIFIED STUDY DEFINITIONS MODEL(USDM)

The Unified Study Definitions Model (USDM) is a framework developed by the Clinical Data Interchange Standards Consortium (CDISC), based on the concept of data interoperability between various systems. It provides a common language for describing clinical trial data, facilitating sharing and reuse of trial data across studies and organizations. The model is based on the Protocol Representation Model (PRM), aligning data exchange with the Study Design Model (SDM), and the Operational Data Model (ODM) from CDISC. USDM is also interoperable with HL7 Fast Healthcare Interoperable Resources (FHIR) standards, supporting data exchange. It promotes the use of controlled terminologies and standard vocabularies to ensure consistency and interoperability of trial data across different systems.

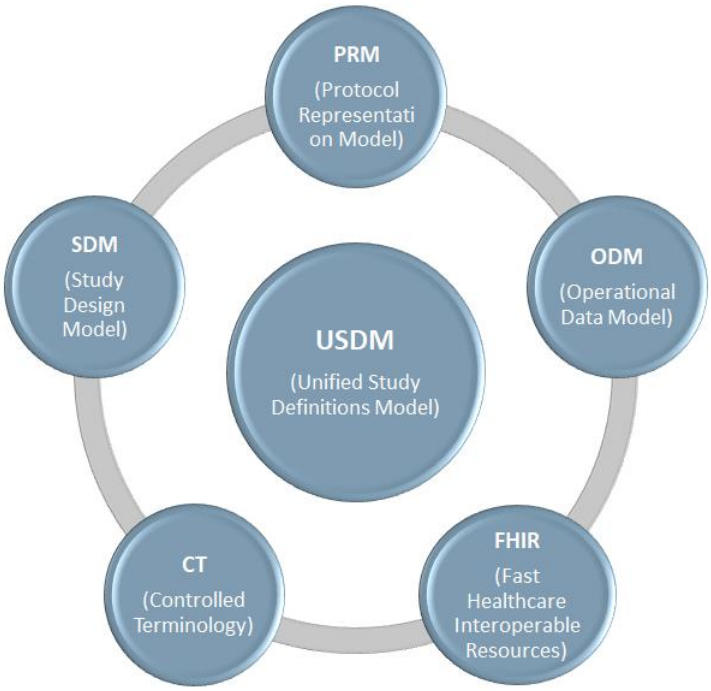

Figure 4. Standards Collaboration

Many of the data elements in USDM can be supported through extended controlled terminology and reference data maintained in a metadata repository. Technology providers can leverage the relationships defined in the model to track amendments, analyze the impact on systems, and quickly push updated information.
USDM is designed to describe study design elements, their attributes, and the relationships between them. It can be primarily classified into:

2.1.1. STUDY DETAILS

Study details comprise the information that helps describe the study characteristics. It primarily contains the following data elements:

- **Study Title** - The study title is a brief description of the clinical trial that summarizes the main purpose of the study. It is usually a few words or a short phrase that accurately and concisely describes the study.

- **Study Description** - It is a summary of the clinical trial that provides an overview of the study design, objectives, procedures, and outcomes. The study description is typically included in the protocol for the clinical trial.

- **Study Version** - It helps to ensure that everyone involved in the clinical trial is using the same version of the study documents. This is critical for maintaining consistency and accuracy in the study procedures and data collection, and for ensuring that the study is conducted by following the approved protocol and any applicable regulations or guidelines.

- **Study Phase** - This describes the stage of development of the clinical trial, such as Phase I, Phase II, Phase III, or Phase IV.

- **Study Type** - The study type is a component of the study details for clinical trials that refers to the type of study being conducted.

- **Protocol Title** - The protocol title typically provides a concise and descriptive summary of the clinical trial and should reflect the primary objective and design of the study.

- **Methodology** - it provides a detailed description of how the study was conducted and serves as a reference for researchers and regulators evaluating the study's scientific validity and ethical integrity.

- **Analysis Population** - This includes information about the targeted study population, such as the demographics, and any other relevant characteristics.

- **Study Sponsor** - The study sponsor details typically include information such as the name of the sponsor organization or individual, the contact information for the sponsor, and any relevant disclosures or conflicts of interest.

- **Approval Regulatory Agencies** - Regulatory agencies are government agencies that are responsible for ensuring that clinical trials are conducted in a safe, ethical, and scientifically valid manner. A few examples of regulatory agencies are:
    - US Food and Drug Administration (FDA)
    - European Medicines Agency (EMA)
    - Pharmaceuticals and Medical Devices Agency (PMDA) in Japan

2.1.2. STUDY OBJECTIVE

This defines the reason or aim of conducting the clinical trial in terms of what treatment or preventive modality would be achieved which can be supported by the analysis of data collected during the study.
To digitize the study objective each objective should be classified as primary and secondary objectives as well as should be attributed with the objective endpoint. The endpoints should be supported with sponsor-controlled terminology containing the description, purpose, and specific level.

2.1.3. INDICATIONS:

Defines the targeted disease or condition under study expected to be treated by the drug to evaluate the drug's effectiveness. Indication code and description are recommended to be governed by a standardized controlled Terminology maintained in the metadata repository by the tech providers to feed to the protocol authoring tool.

2.1.4. INCLUSION/EXCLUSION CRITERIA(I/E):

This defines the eligibility criteria of the characteristics or requirements applied to a potential/study subject like age, gender, certain medical conditions, etc. to decide whether he will participate in the study or not. These can be supported by a controlled terminology for standards set of definitions.

2.1.5. SCHEDULE OF ACTIVITIES:

This includes planned visits or time points at which the treatment or assessment is done. It comprises a detailed activity plan defining each planned activity to be done at a specific time based on certain

predefined conditions. Currently, the Schedule of activities is captured in the form of a table supported by conditions and exceptions mentioned as notes.

VISIT	Screening	Cycle 1 Day 1	Cycle 1 Day 8	Cycle 2 Day 1	Cycle 2 Day 8	Cycle 3 Day 1	Cycle 3 Day 8	Cycle 4 Day 1	Cycle 4 Day 8	Cycle 5 Day 1	Follow up
ASSESSMENTS											
Infomed Consent	X										
Screening	X										
Demographics	X										
Hematology	X		X		X		X		X		X
Vital Signs	X	X		X		X		X		X	
Drug Dispensing		X	X	X	X	X	X	X	X	X	
Biochemistry	X										X

Table 1. Schedule of Activities

To digitize the Schedule of Activities the Visit Structure and activity need to be standardized as reference data where each visit and activity have a standardized description, and the attributes to support the study workflow. This means that each visit needs to be marked with the start and finish conditions which could be timing, activity along with other supportive descriptors like mode of contact, Place of Visit, etc.
Technology providers can process the lineage between the Schedule of Activities, visit information, and data element definition for clinical data collection in a Metadata Repository. Thus, it can be used for automation of study build in EDC as well auto-generation of the Data Transfer Specifications for Labs, eCOA, ePRO, etc.

2.1.6. STUDY WORKFLOW

Study workflow is an elaboration of the Schedule of Activities. It connects all the activities including the assessments and procedures at each planned visit defined by how a visit is transitioned from and to another visit. the next and previous visits or activities within a visit along with the condition and criteria to trigger these transitions. The study workflow is further supported by the intercurrent events (like abnormal lab results, Visit delays, missed doses, and adverse events) that are not a part of the happy study path. These events are defined along with the strategy of actions to be taken when encountered.

2.2. APPLICATION PROGRAMMING INTERFACE (API)

The application programming interface (API) serves as a bridge between multiple systems that adhere to data standards. APIs act as a fundamental component of the integration framework between the protocol authoring system and the consuming system, based on the data flow section required. The ability to transfer data to multiple systems simultaneously ensures parallel execution of study set-up activities with consistent information. The defined mapping also allows for maintaining traceability between the protocol and clinical trial systems, facilitating quick analysis of impacts and prompt action-taking.

In providing out-of-the-box APIs, product vendors and organizations play a crucial role. The APIs' compliance with data standards allows them to support interoperability between models via auto-transformation of data from one model to another, leveraging the relationships defined by CDISC. This kind of technical work is not new for technloy and can be achived in automated ways. Product vendors and organizations are instrumental in providing pre-built APIs that comply with data standards and enable interoperability between different models. By leveraging the relationships defined by CDISC, data can be automatically transformed from one model to another. While this kind of technical work is not novel, it can be accomplished using automated methods.

2.3. CONTROLLED TERMINOLOGY (CT)

Controlled Terminology (CT) refers to a standardized list of values that facilitates the selection of data elements within the Unified Study Definitions Model (USDM) in clinical trials. The CT is a combination of CDISC-provided terminology as well as terminology defined by the study sponsor.

Currently, many technology providers and product vendors use CTs as part of their metadata repository solutions for data collection. This existing infrastructure can be expanded to support study design data elements, allowing for the incorporation of standardized terminology into the design and execution of clinical trials.

3. DOWNSTREAM SYSTEMS

In the realm of clinical trials, various downstream systems (such as EDC, eTMF, CTMS, IWRS/IVRS) are commonly employed. To ensure maximum flexibility, DDF has been designed to be both product and sponsor-agnostic. This means that any product or platform capable of consuming data in a CDISC standard format can be easily integrated into the digital data flow.

HOW DOES DDF AIM FOR DIGITAL TRANSFORMATION?

DIGITAL PROTOCOLS

The aim is to move from a document paradigm to a digital format of the clinical protocol. Technology can be used to collect and combine the study element to create digital protocols to avoid the traditional way of paper-based documentation the protocol information. With the use of technology, study elements can be collected and combined to create digital protocols, which can replace traditional paper-based protocols. Digital protocols provide numerous benefits, such as improved efficiency, accuracy, and transparency.

CONNECTIVITY OF DATA AND PROCESSES

One of the primary objectives of the DDF initiative is to establish seamless connectivity between data and processes. This involves leveraging the platform capabilities of technology providers to create end-to-end lineage and automated information flows using APIs and metadata mappings. By ensuring that the right data is available at the right time, the DDF provides greater oversight to all processes, enabling better decision-making and improved outcomes.

ADVANCE ANALYTICS

By digitizing the clinical protocol and establishing its lineage to downstream systems, DDF creates opportunities for advanced analytics to optimize the study design and improve analysis population, scheduling of activities, patient engagement, operational efficiencies, and compliance. With access to a unified and standardized data flow, stakeholders can leverage advanced analytics tools to gain insights that were previously difficult to obtain, leading to better decision-making and improved clinical outcomes.

OPEN AND FLEXIBLE SOLUTION

DDF provides an open and flexible solution by adopting a vendor-agnostic approach, which enables the creation of customized solutions based on the strengths and preferences of each platform or solution vendor. This allows for easy scaling for new features, adoption of changes, and collaboration on outcomes in a plug-and-play mode.

DEVELOPMENT PRINCIPALS

The development principle of DDF is to promote the interoperability and efficiency of clinical trials by establishing a common language for data exchange. By adopting DDF standards, clinical trial data can be easily shared and reused across different studies and organizations, which can save time and reduce costs. Digital Data Flow (DDF) is a set of CDISC data standards used to facilitate the exchange of clinical trial data between different organizations and systems.

OPEN-SOURCE

Open-source technologies play an essential role in clinical trial solutions. Open-source software has several advantages, including cost-effectiveness, flexibility, and the ability to collaborate and share knowledge. ODM-XML Converter is an open-source tool that converts clinical trial data between the Operational Data Model and the XML format, enabling the interoperability of clinical data across different

systems. Having flexibility with open-source technology can help technology providers to use cost-effective software and make it available at a decent cost to customers. The use of open-source technologies helps to promote transparency, collaboration, and innovation in clinical research.

VENDOR AGNOSTIC

Digital Data Flow (DDF) is a vendor-agnostic initiative for clinical trials. By being vendor-agnostic, DDF promotes interoperability between different systems and technologies, in a plug – and - play manner. The use of APIs keeps the systems integrated rather than making them tightly coupled with each other. The data mapping between systems can utilize these flexibilities to create vendor-agnostic clinical computing platforms.

AGILE DEVELOPMENT

Agile development is an iterative and collaborative approach to software development that emphasizes flexibility and responsiveness to changing requirements. DDF supports Agile development by providing a flexible framework for data exchange that can be adapted to different systems and technologies. DDF also supports Agile development by promoting collaboration between different stakeholders involved in clinical trials. This is the era of agile for software development and being DDF agile supportive will help organizations to create software in their agile methodology.

DYNAMIC ALIGNMENT TO STANDARDS

The DDF initiative supports dynamic alignment to standards by providing a flexible framework for data exchange that can be adapted to different systems and technologies while still maintaining adherence to CDISC standards. The guidelines give solution providers an opportunity to ensure that data is mapped and transformed correctly from one system to another when the systems follow a data model complaint to the CDISC standards. However, the solution providers have to ensure that the data model is flexible enough to absorb the change in standards as they evolve to minimize the impact on ongoing trials and reduce the change management cost.

SUMMARY

Technology can impact DDF in several ways. Technology can enable the creation of digital protocols, which can simplify and streamline the clinical trial process. Technology can provide connectivity between different systems, enabling the flow of information between them and creating an end-to-end lineage of data. Technology can facilitate the use of controlled terminology, ensuring consistency and interoperability of trial data across different systems. Finally, technology can provide an open and flexible solution, allowing different platforms and vendors to easily scale and collaborate.

REFERENCES

https://en.wikipedia.org/wiki/SDTM
https://www.cdisc.org/ddf
https://www.transceleratebiopharmainc.com/initiatives/digital-data-flow/

CONTACT INFORMATION

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

Prasoon Sangwan
prasoon.sangwan@tcs.com
www.tcs.com

Piyush Singh
piyushkumar.singh@tcs.com
www.tcs.com