

## End-to-end data and metadata: It's not only a dream, it's a reality

Gérald NEVEU, Danone Nutricia Research, Palaiseau, France

### ABSTRACT

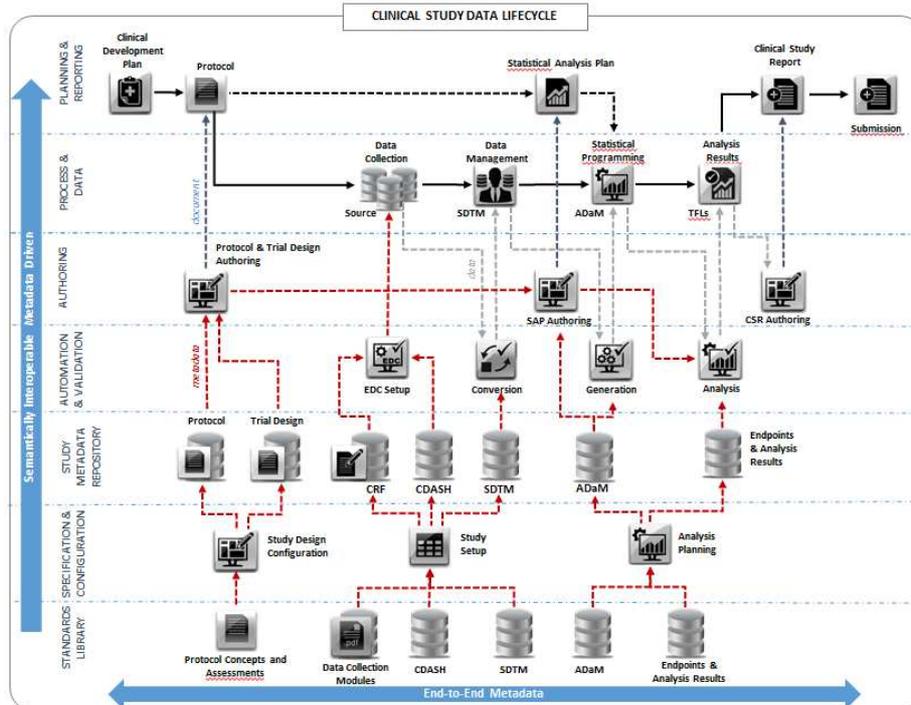
The landscape of standards (i.e. CDISC ODM, CDASH, SDTM, AdAM etc.) and available off-the-shelf (OTS) eClinical and data visualisation tools is mature and ready to implement a coherent end-to-end data and metadata strategy for clinical research – it's not only a dream, it's a reality. We have communicated in the past on how we've pieced together this End-to-End process. This presentation will show how standardization combined with a data visualisation tool can be leveraged to seamlessly go back-and-forth through the end-to-end process: From CSR, back to AdAM results and metadata, back to SDTM and the CDASH forms, and the Trial Design metadata.

### INTRODUCTION

This paper will focus on the main aspects on how Danone Nutricia Research imagines and builds an End-to-End process based on CDISC standards and how a Data Visualisation tool can be plugged on it in order to visualize efficiently and easily your data and metadata from the protocol to the CSR narrating your Clinical Study stories.

### A NEW VISION OFFERING A NEW START

Five years ago, Danone was in prehistoric ages processing Clinical Data based on a paper CRF process and outsourcing all Data Management Activities with a low level of standards. The time came to change this state and renovate the Clinical Data Strategy in adopting new generation of tools and adopting a 100% CDISC standards based process taking advantages of all their benefits. During the last PhUSE SDE event in February 2016 at Utrecht, Peter Van Reusel from Business & Decision Life Sciences© presents a slide (figure 1) summarizing the Danone's vision of an End-to-End process enabling to link each component of a Clinical Study Data Lifecycle from the Protocol to the CSR.



Restricted © Business & Decision Life Sciences 2016 All rights reserved.  
 Figure 1: Clinical Study Data Lifecycle, Paul Van Reusel, Business & Decision Life Sciences©

## PhUSE 2016

On the paper, this vision seems to be a dream. Indeed, control an entire flow like this would request a huge workload multiplying the use of different tools, asking for several operational skills in different domains, many resources with different profiles. That's without taking into account on use of CDISC standards.

At this time, Danone Nutricia Research completed the implementation of a flow controlling the Clinical Data part from the Clinical Data collection to their storage in a Data Repository including the design.xml and .xpt files necessary for a submission for each study totally designed in SDTM. The name of this flow is DanFlow.

This Data Repository containing as well as Clinical Data but also Study documents: a Metadata Catalog was built enclosing all metadata related to all studies managed by the company reinforcing the Danone's Legacy. This catalog is a database structured on CDISC SDTM Trial Design Tables standards.

To complete this flow, Danone Nutricia Research is working on the statistical part with statistical programming (AdAM) and analysis result (TFL) still under implementation and same for Data Visualisation templates.

It's time to come "back to the future" and discover this End-to-End process in details and how a Data Visualisation tool can emphasize the data discovery of this huge source of information.

### CLINICAL DATA FLOW

Five years in the past, after different benchmarks and discussions with some providers, the new vision was more concrete and focused essentially on clinical data management part from the data collection to the storage at the beginning. The DanFlow was then created integrating four different layers: all based on CDISC tools and connected by an ODM file enabling the transmission of clinical metadata and data (figure 2).

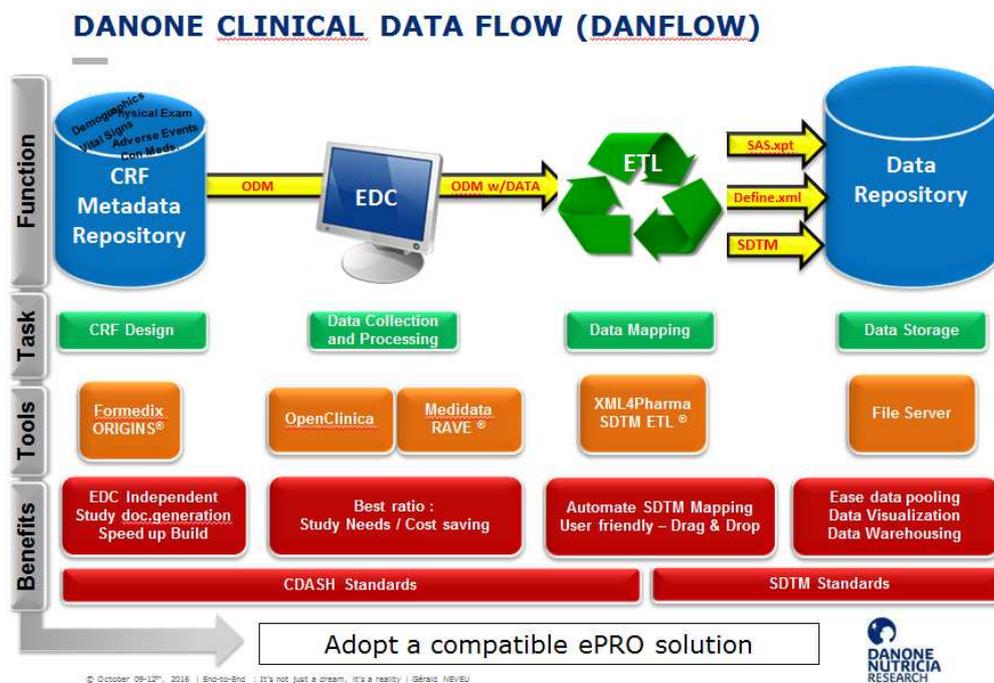


Figure 2: DanFlow and his four layers

The first layer encloses the whole Metadata Repository with forms, items, edit checks and codelists. The tool used is Formedix© Origin®, a CDISC certified tool using the ODM markup to hierarchize these CRF elements. They are designed in CDASH by Danone Nutricia Research team and used to design an eCRF outside from the EDC system itself. An ODM file is then created including CRF specifications.

The second layer is the EDC system and thanks to the use of an ODM file, we have the possibility to use any ODM-compliant EDC system. We can then choose the system based on study complexity and have a better Study Needs/Cost saving ratio. Two EDC systems are used by Danone Nutricia Research. From this EDC, an ODM can be generated enclosing the CRF metadata but also clinical data.

One of the biggest challenges left: how to transform efficiently this CDASH data in SDTM? On the CDISC website, a list of CDISC certified tool contains an ETL scripted by XML4PHARMA©. It is the third layer. As this tool is CDISC certified, it contains some presets to map the data easily in SDTM with a drag and drop environment after having mounted a define.xml file enclosing the CDISC SDTM library. From your CDASH data in the ODM file, the SDTM ETL® tool will then transform them in SDTM structure in define.xml and xpt study-specific files.

These files will be then stored inside the fourth layer: a file server able to store the Data Repository with some study documents. The exploitation of a SDTM data repository has several advantages among which capitalize the data in a Data Warehouse by pooling data more easily and plug a data visualisation tool on it.

## PhUSE 2016

An ePRO solution has been integrated to this data flow in order to profit of all these gains.

### CLINICAL METADATA CATALOG

The second step of this new Clinical Data Strategy is focused on the creation of a Metadata Catalog. Once again, CDISC standards were a good opportunity to start thinking about which information need to be captured in this Clinical Metadata Catalog.

Based on source of information we have and the DanFlow, another data flow was created (figure 3).

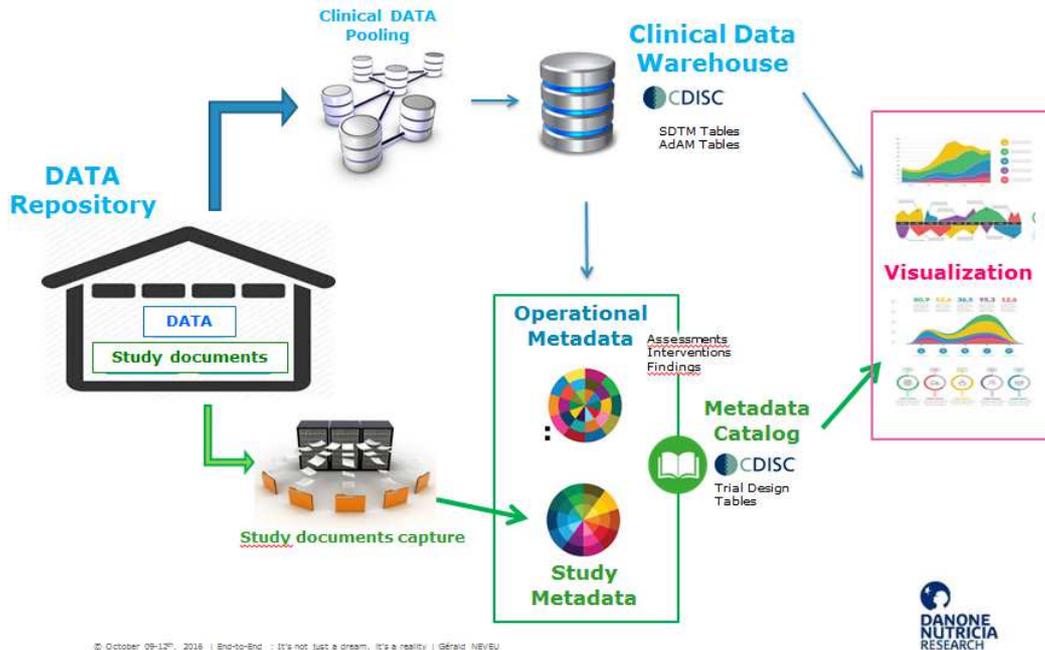


Figure 3: Data Flow for Metadata Catalog creation

For each study, some study documents were imported into the Data Repository and will be the source to complete the CDISC Trial Design Tables requirements. However, CDISC offers enough flexibility in term of information thanks to the Trial Summary table. And then, after several working group with the Clinical stakeholders from Study Core team, essentials information to be included in this table were defined.

To automate the process, an excel file template was created to integrate these essentials information from each study documents for each study. Once completed, all these excel file information are aggregated into a CDISC SDTM database based on Trial Design Tables standards. For future studies, the PRM (Protocol Representation Model) could be used to aggregate semi-automatically these information in the existing Metadata Catalog.

Some clinical metadata can be extracted from Clinical Data such as but not limited to: assessments, interventions and other study design specifications more focused on operational aspects. This can be done easily from Clinical Data Warehouse which has been created resulting from the encapsulation of Clinical Studies SDTM Databases generated by the DanFlow.

It shows a continuum between these both flows as specified in the Peter Van Reusel's slide: all elements from the protocol to the clinical data are now compiled in CDISC databases.

But is it enough? We can go further by adding clinical operations metadata to this clinical data and metadata.

### CLINICAL OPERATIONS METADATA

Outside of this vision, another topic is interesting to be implemented in this vision. Clinical operations metadata defined by Source Data Verification, Queries, Time to Data Entry, Data changes and other multiple items...

As a reminder, all these data are also standardized in CDASH and then the same kind of structure is used across studies and even the ODM file exported from the EDC is the same. So, we have the possibility to build a common flow to capture this data and exploit them (figure 4).

For this, we will export data from each EDC system by using their WebServices: a command will be sent to the server and data will be sent back as an answer to the request. We can then export these data per batch and on-the-fly from the system and use an ETL tool to transform this data for exploitation.

# PhUSE 2016

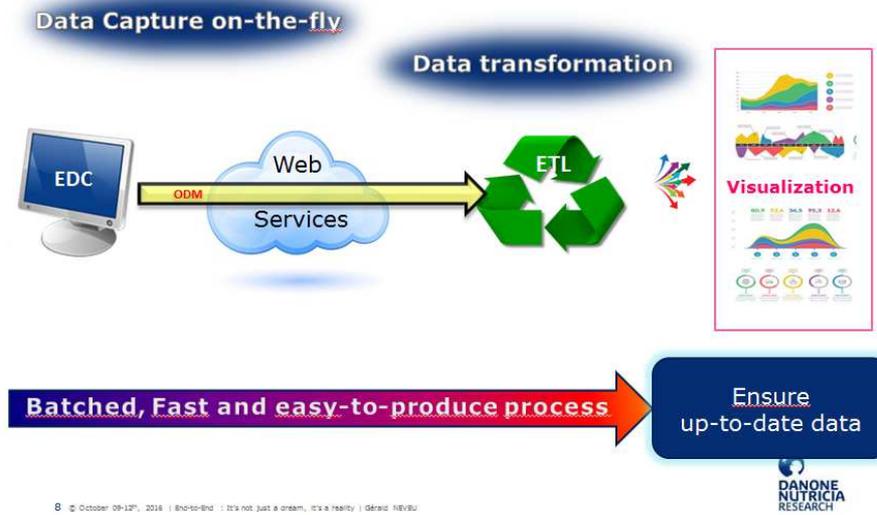


Figure 4: Clinical Operation Metadata generation flow

At Danone Nutricia Research, we use an ETL tool endorsing the step-by-step automatization of this flow from the EDC to the transformed database per batch or on-demand. The flow needs to be executed quickly in order to ensure the most up-to-date data. Indeed, during the course of a study, if we need to monitor a situation with a risk then the data needs to be refreshed continuously to react appropriately and hurriedly.

## DATA VISUALISATION

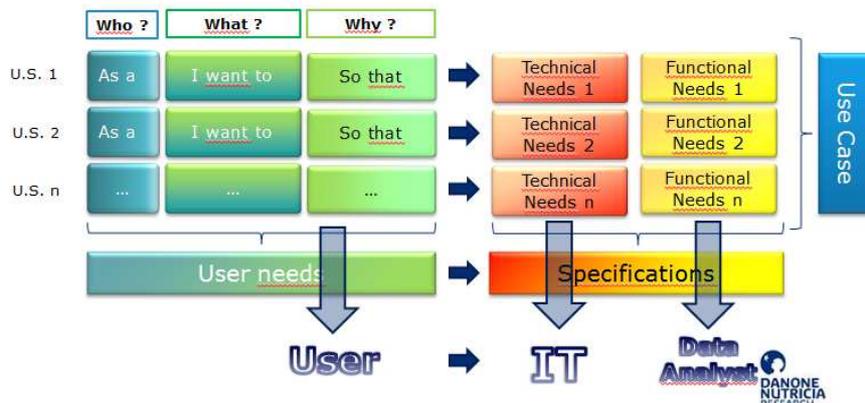
The Data Visualisation offers the opportunity to look at a big volume or/and diversity of data easily and quickly accelerating the decision-making process. For this, some data visualisation templates need to be defined and so capture the user needs efficiently.

The Agile methodology often used in Business Intelligence domain is a perfect model. Each technical (IT systems if needed) and functional (software feature, visualization) need has been translated from user story elements defining a Use Case: “Visualize clinical data from the protocol to the Database Lock” (figure 5).

For each end-user profile, defined here by each clinical core team member or clinical stakeholder, functional needs will be translated into a data visualisation template.

- Users Story definition

- Common language to describe what a user does or needs to complete his job and be the support to define the list of technical needs for IT and functionalities to be designed by the Data Analyst.



© Confidential | 04 FEB 2016 | Implementation of clinical data and metadata dashboards - from User Story to End Usage | G. NEVEU

Figure 5: From user needs to specifications.

## FROM A DREAM... TO A REALITY

## PhUSE 2016

Three databases have been identified and produced from an End-to-End process. Associated to this, some data visualization templates have been designed in order to meet with requirements from user stories. The purpose is to give the maximum of information for each study from the protocol to the database lock and additionally clinical operations data to the adequate end-user profile. At the beginning, this vision was perceived as a dream but now it becomes a reality from which many benefits can be exploited.

### USE CASE & BENEFITS

Each database or source of information has a purpose such as:

- ⇒ Metadata Catalog permits scientists to have a better overview on what has been done in the past. It help to develop robust Clinical development plan and find the best design for a future study based on the success of previous ones. This Metadata Catalog helps biostatistician to build a statistical plan and find adequate Clinical studies to make a Meta-analysis based on past studies and their outcomes. It is also a chance for managers to have global KPIs and study specific KPIs in term of budget, timelines...
- ⇒ The Clinical Data Warehouse is a master piece in term of data capitalization and cross studies data exploitation. It helps the biostatistician team to standardize their analysis programs due to the fact that the structure is the same across studies. Additionally, this source of information can be pointed out by a Data Visualisation tool to animate Data Review Meeting during a course of a study by producing medical review listing, patient profile.
- ⇒ The Clinical Operations data, extracted on-the-fly from each EDC system, are used to produce Clinical dashboards and Data Management Metrics. Dashboards help a Study Core Team to better anticipate any emergency and anticipate efficiently each operational constraint during the course of a study. For any company engaged in a Risk-based Monitoring strategy this can help also to lead this strategy and readapt it. Data Management Metrics contains some KPIs elements related to Data Entry, queries management, Data Changes, support to seek problems or misunderstanding from sites on Clinical Data themselves.



All these different components constitute a storyboard describing each clinical study from the protocol to the database lock with operational parameters. But this journey is not finished: a building zone is opened to integrate analysis, TFL and CSR part in this workflow. It will complete this End-to-End process and will allow us to finalize this vision and append new elements from the database lock to the CSR.

### RETURN-ON-EXPERIENCE

To make it real, the idea of adopting CDISC standards all along this process is a key element and nothing would be real on today if it was not the case. Without these standards, the work would have taken more time because...you would have need to create your own standards at the end before beginning to work on the concretization of this vision.

CDISC standards are efficient rules defined by many companies all over the world with experts, technician, KOL and for sure the work done by this instance is helpful for all people in relation with Clinical Data. The standardization is itself an element enabling the automatization and the creation of an End-to-End process and make it more efficient. But, as CDISC is totally designed to manage clinical data, it makes it easier because it gives some guidelines to follow on which kind of data needs to be collected.

By using these standards, each element in this End-to-End process are compatible and a link between them can be done easily, data encapsulation is easy to manage. A library of scripts, CRF forms, and visualisation templates can be created to automate the job producing clinical metadata and data as described in the Paul Van Reusel slide. Once generated, these sources can be targeted by a Data Visualisation tool.

Design of visualisation is easy to manage but their specifications are more complicated to write. Indeed, as a lot of different stakeholder will use them, the starting point is to listen them and transform their needs in a reality. To do that, a robust methodology needs to be set up, that's why Danone Nutricia Research decided to use the Agile methodology who demonstrates already his efficiency in this domain with Business Intelligence applications since few years now.

At Danone Nutricia Research, use cases were find out through the results of applying an agile methodology with each stakeholder defining their needs based on their daily job tasks.

## PhUSE 2016

### REFERENCES

PhUSE SDE Event, February 4<sup>th</sup>, 2016, Utrecht. “How a visual representation can help us understand the potential of clinical metadata”, Peter Van Reusel, Business & Decision Life Science

PhUSE SDE Event, February 4<sup>th</sup>, 2016, Utrecht. “Implementation of clinical data and metadata dashboards – from User Story to End Usage”, Gérald NEVEU, Danone Nutricia Research

List of CDISC ODM certified products, <https://www.cdisc.org/resources/odm>

CDISC Protocol Representation Model, <https://www.cdisc.org/standards/foundational/protocol>

SDTM IG, <https://www.cdisc.org/standards/foundational/sdtm>

### CONTACT INFORMATION

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

Gérald NEVEU

Danone Nutricia Research

RD128 – Avenue de la Vauve

91767 Palaiseau Cedex FRANCE

Work Phone: +33 7 77 44 15 41

Email: [gerald.neveu@danone.com](mailto:gerald.neveu@danone.com)

Brand and product names are trademarks of their respective companies.