PhUSE 2010

Paper CD12

# Bridging the Communication Gap
# Implementing Metadata Standards in Trial Protocols

Annabel Acs, Novo Nordisk, Søborg, Denmark
Pia Hjulskov Kristensen, Novo Nordisk, Søborg, Denmark

**ABSTRACT**
Like many other companies, Novo Nordisk has recently invested a lot of time and resources towards standardising how data is collected and analyzed. When facing the challenge and seemingly uphill battle of standardisation, it has been important to consider three factors. First and foremost, one must consider the data standards at hand; second the type of technology that will support the standards; and finally the process by which the standards are maintained. Novo Nordisk has recently developed a metadata driven Clinical Data Warehouse (CDW), which is used to implement data and programming standards based on CDISC standards. After implementing the technology, it has proven even harder to implement the work processes that will support the new technology and standards. In an effort to create a process around maintaining clinical data standards before they reach the clinical data warehouse, an Excel-based tool has been developed, which supports the use of CDISC and Novo Nordisk metadata standards at the protocol, CRF and database level. Aside from this tool being relatively inexpensive to develop, it is also an easy to use, low tech solution, making it the perfect tool for Medical Directors, Trial Managers, Data Managers, Global Librarians and Biostatisticians.

**INTRODUCTION**
Imagine the fire triangle, which is a model used to illustrate the three essential ingredients required to build a fire: heat, fuel and oxygen. If one of these elements is missing then it is simply impossible to build a fire.

In the same sense, it is difficult to establish clinical data standards without establishing three important elements: standards, technology and processes. Theoretically, any effort to implement clinical data standards would crumble without all three of these factors fully developed. However, unlike the fire triangle, where all three elements should be equally present at the same time, the three elements in the 'standards circle' (see figure 1) should be implemented incrementally. In other words, before beginning to develop the right technology, one must first establish and understand the standards to be used and how the technology will support these standards. In the same sense, before fully implementing a functioning process, the technology must be up and running.

On another level, one could say that each layer of the standards circle becomes increasingly complex and more difficult to modify. With the first layer, you are working with codes, programs, and data models – all very complex, but perhaps easier to modify than the third level where are you are dealing with people, communication, politics, and opinions.

The focus of this paper is to discuss how these three factors were applied to implement a process for using protocol metadata standards.

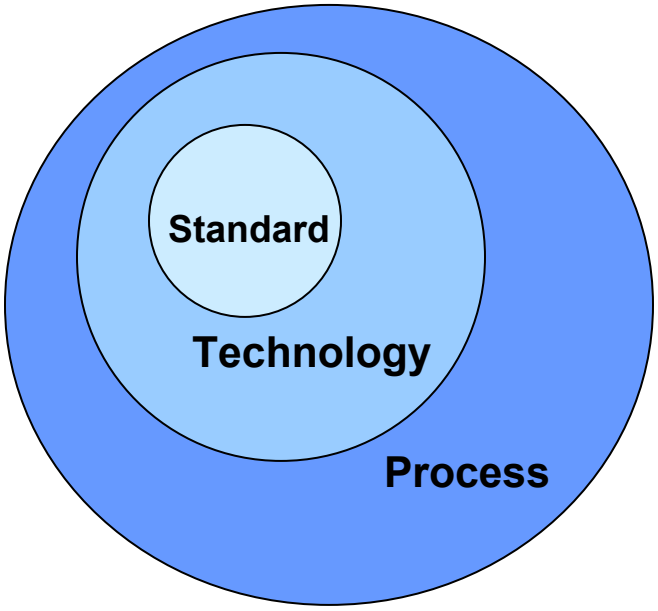



FIGURE 1

**CLINICAL DATA WAREHOUSE AND NOVO NORDISK**

In the world of clinical trials at Novo Nordisk, standards have been historically created in silos. Trial managers create their own protocol standards, Data managers follow their own database standards and statistical programmers each create an individual programming standard, on top of that, diabetes projects create their own standard while haemophilia projects create another. Unfortunately by working in silos; a lot of extra work is created to certain colleagues and NN is ultimately prevented from fully utilizing all the clinical data that has been collected throughout the years.

It therefore goes without saying that, like most other pharmaceutical companies there was a need to standardise clinical data for collection and analysis. At Novo Nordisk, the Clinical Data Warehouse, which is a metadata driven warehouse, has been developed to support the development of standard statistical programs for data analysis. The system primarily supports the development of an internal Novo Nordisk data standard; however, external standards such as CDISC have also been used to define the standard libraries.

The decision to develop a system around CDISC standards benefits the company by providing a consistent standard for metadata and a data model. However it also will support future submissions to the FDA and possible data sharing with other health organisations.

The Clinical Data Warehouse, which is a metadata driven system, was created to support to generation of standard statistical programs. By collecting and organising data based on CDISC SDTM terminology and a standard data model, it is possible to develop standard programs that can be used from one trial to the next.

Within the CDW, metadata codes have been defined further defined as either clinical or trial metadata. Clinical metadata can be thought of as a general pool of metadata that includes topic codes and generic trial designs. From the clinical metadata it is possible to select trial metadata, which describes the data that will be collected for a specific trial. For example, the trial metadata could include the visit structure, trial flowchart, and trial description.

The metadata standards for CDW are primarily based on CDISC SDTM standards. For example similar to

CDISC standards, the codes in the system are categorised by Topic codes, which includes Findings, Interventions, Events and Supplemental Qualifiers, and codelist codes. The naming conventions for the majority of these codes are based on CDISC SDTM terminology, however if a CDISC standard has not been available then an internal NN standard has been developed.

Despite the workload required to implement clinical metadata standards in the CDW, the benefits will allow Novo to work more efficiently and take on even more trials. First of all, it will be possible to develop standard statistical programs, which in the future should allow programmers to use universal programs for most trials and free up resources to develop more complex programs. It should also be less time consuming and more straightforward to conduct cross trial analysis since less time will be spent aligning and locating data. Furthermore, it will be possible to flag incorrect data using standard clinical; for example, it is possible to set possible lab ranges for findings in order to catch outliers. Finally, standard metadata can facilitate later queries. For example, if there is a need to look at a specific trial – perhaps due to a request from the health authorities – then it is possible to conduct a keyword search on trial descriptions and find the needed trial. In the past, it was very difficult to search for old trials because one would have to physically look through old protocols.

After the standard and supporting technology have been established then it is possible to develop, implement and test a process. The primary end-users of the CDW are statisticians and programmers who have developed processes around entering metadata and implementing the use of standard programs for statistical analysis. CDW also demands cooperation from data managers who need to derive codes in OC and map from OC to CDW. However, by default data managers are involved with data standardisation and CIDSC therefore it was relatively straight forward to incorporate them into the process.

On the other hand, the greater challenge was to create a working process between trial management and statistics that would ensure standards are used between the protocol, database and CDW. The focus of this paper is therefore to discuss how the standards, technology and process have been optimised to ensure a connection between the protocol development and CDW (see figure 2).

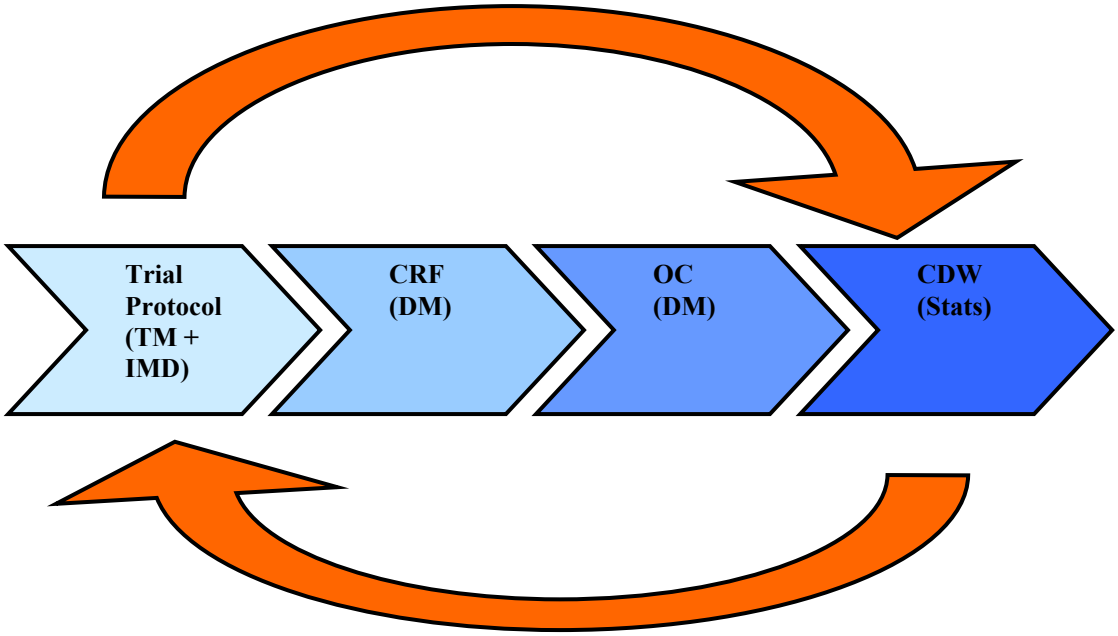



FIGURE 2

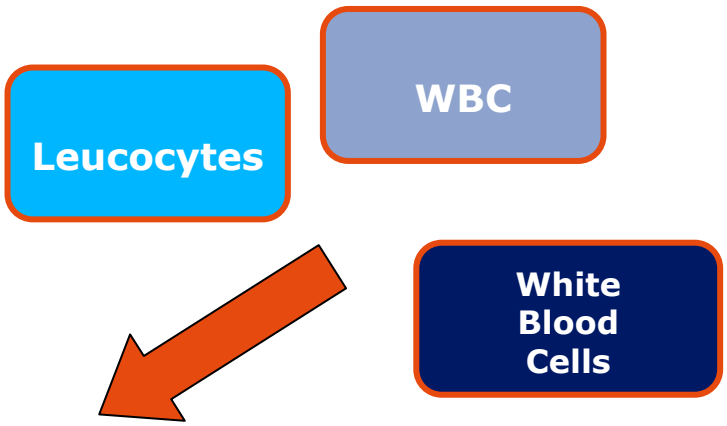
**THE PROTOCOL METADATA TEMPLATE AT NOVO NORDISK**

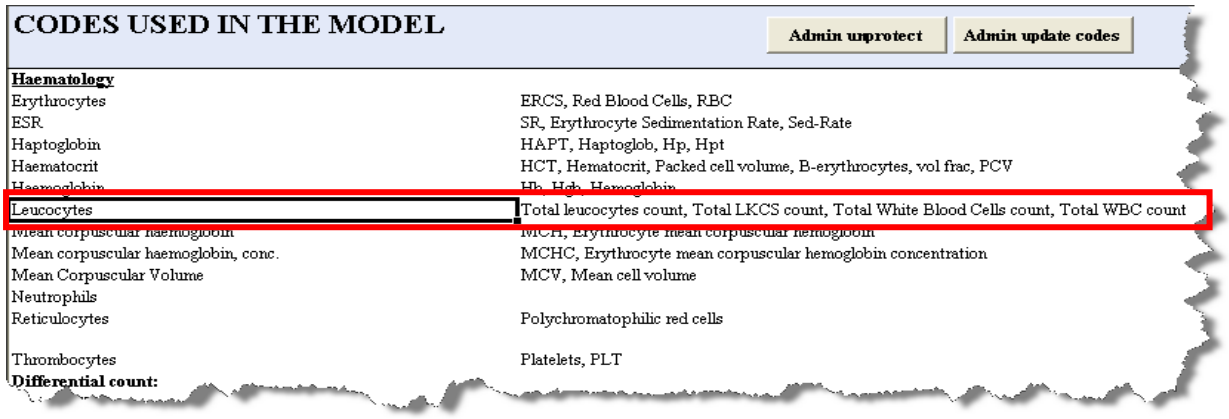
Before the Protocol Metadata Template, there was little standardisation across protocols, so for example to describe leucocytes, one protocol might use 'WBC' or 'white blood cells' (see figure 3). The argument for doing this from a trial management perspective was that the protocol is written to communicate with investigators who in different countries prefer different terms. However, since it is required that the CRF is based on the protocol, all of the non-standard terms used in the protocols would appear in the CRF and ultimately the database. In the long run this causes extra work for programmers and statisticians who need to align the data for analysis. In the interest of saving time for statisticians and programmers, it makes sense to prevent the use of non-standard terms in the protocol and CRF by creating a tool that aids the use of standard terms in all protocols.

Once the CDW was developed and it was established that CDISC standards would be primarily followed, then it was a matter of developing a tool that would support the standards used in CDW. The solution to this was the protocol metadata template, which was developed in collaboration between a trial manager, a programmer and an OC specialist. Therefore both the CDISC standards used in CDW as well as the requirements of the protocol were considered. The standards used in the protocol metadata template were primarily based on CDISC standards (see Figure 3), for example using leucocytes as a standard lab code. In cases where CDISC standards are not available, it was necessary to develop an internal standard.

In order to ensure that there is a continuous evaluation of new standards, trial managers are required to submit new metadata code requests to a standards committee, which is responsible for evaluating and implementing codes. This process is especially relevant for new therapeutic areas, where several protocols are under development and a new standard needs to be established. In general, it has been a benefit to Novo Nordisk that our primary therapeutic area is diabetes, which is already a highly researched area and therefore there are more external standards available. However for less researched therapeutic area where there is almost no external standard, it is necessary to have a committee assigned to setting an internal standard.

In order to fully realise the benefits of the CDW, it is important to ensure that metadata is consistently throughout the entire data collection and analysis process. The protocol metadata template standardises protocol terminology in reference to the clinical and trial metadata defined in CDW. For example, the same standard trial metadata used for a trial description in the CDW and is also used for the trial description in the protocol.



CODES USED IN THE MODEL

Admin unprotect | Admin update codes

| | |
|---|---|
| **Haematology** | |
| Erythrocytes | ERCS, Red Blood Cells, RBC |
| ESR | SR, Erythrocyte Sedimentation Rate, Sed-Rate |
| Haptoglobin | HAPT, Haptoglob, Hp, Hpt |
| Haematocrit | HCT, Hematocrit, Packed cell volume, B-erythrocytes, vol frac, PCV |
| Haemoglobin | Hb, Hgb, Hemoglobin |
| Leucocytes | Total leucocytes count, Total LKCS count, Total White Blood Cells count, Total WBC count |
| Mean corpuscular haemoglobin | MCH, Erythrocyte mean corpuscular hemoglobin |
| Mean corpuscular haemoglobin, conc. | MCHC, Erythrocyte mean corpuscular hemoglobin concentration |
| Mean Corpuscular Volume | MCV, Mean cell volume |
| Neutrophils | |
| Reticulocytes | Polychromatophilic red cells |
| Thrombocytes | Platelets, PLT |
| **Differential count:** | |

FIGURE 3

**TECHNOLOGY USED FOR THE PROTOCOL METADATA TEMPLATE**

When developing the correct technology to support standards, it is undoubtedly essential to consider the end-users. In the case of protocol metadata standards, the end-users are primarily Trial Managers and Clinical Trial Administrators. In general, people within trial management do not work with databases or programming therefore making it essential to create a system that is familiar and relatively easy to use. Excel was a particularly good option because it is familiar to most people, relatively easy to work with and available on everyone's computer. In addition to being user friendly, Excel was also an inexpensive solution. The only cost associated with the template was spent on an Excel consultant to develop the programming. When considering how much time and money went into developing the CDW, this was a relatively low cost solution.

Apart from Excel, it was also considered that the CDW should include an interface trial managers where could enter metadata. This would have been a much more complex and expensive technological solution, which would require expanding functionalities in CDW and is currently de-scoped. Considering how complex and time consuming is has been to develop the CDW, it was probably a better idea to develop and perfect the technology, standards and processes on a smaller scale. In the meantime, the protocol metadata template has provided a simpler and lower-cost solution to align protocol metadata standards with CDW. In the future, it could serve as a model for how to expand the CDW to include protocol metadata standards.

Excel is also fairly adaptable, making it a good program to use if frequent updates will be needed. Since there was a bit of trial and error involved in the development of standard templates, it was a great benefit that the Excel templates could be easily modified. The template has now been upgraded four times and each time it was possible to assess the needs of the users and modify it accordingly.

**PROCESS FOR USING PROTOCOL METADATA TEMPLATE**

The process built up around the PMT and the standards is arguably the most complex and difficult step to implement simply because it involves people. When the templates were developed, the intention was that they would be filled in while the protocol is written in order to ensure that the visit structure, Trial Design, and flowchart are aligned with the metadata standards in CDW. It therefore was required that Trial Managers fill in the PMT while they are writing a trial protocol (see Figure 4). However, this proved to be a difficult process change to implement mainly because it was a new and rather time consuming task. As a result, the metadata templates were usually left unfinished until the protocol was finalised and then the task of filling them in was delegated to administrative personnel.

Therefore, when trying to streamline and integrate how people work, several roadblocks were encountered. First of all, it is difficult to convince people to take on more work and even more difficult to ask people to spend the time to learn a new tool. Furthermore, if a task appears too administrative then it's difficult to convince people with the right background to take responsibility. One answer to this is to train colleagues and ensure that they fully understand the benefits of the work they're taking on. However, in the first two years of implementing this process it was clear that without more management support, training alone would not get people to use the template as initially intended.

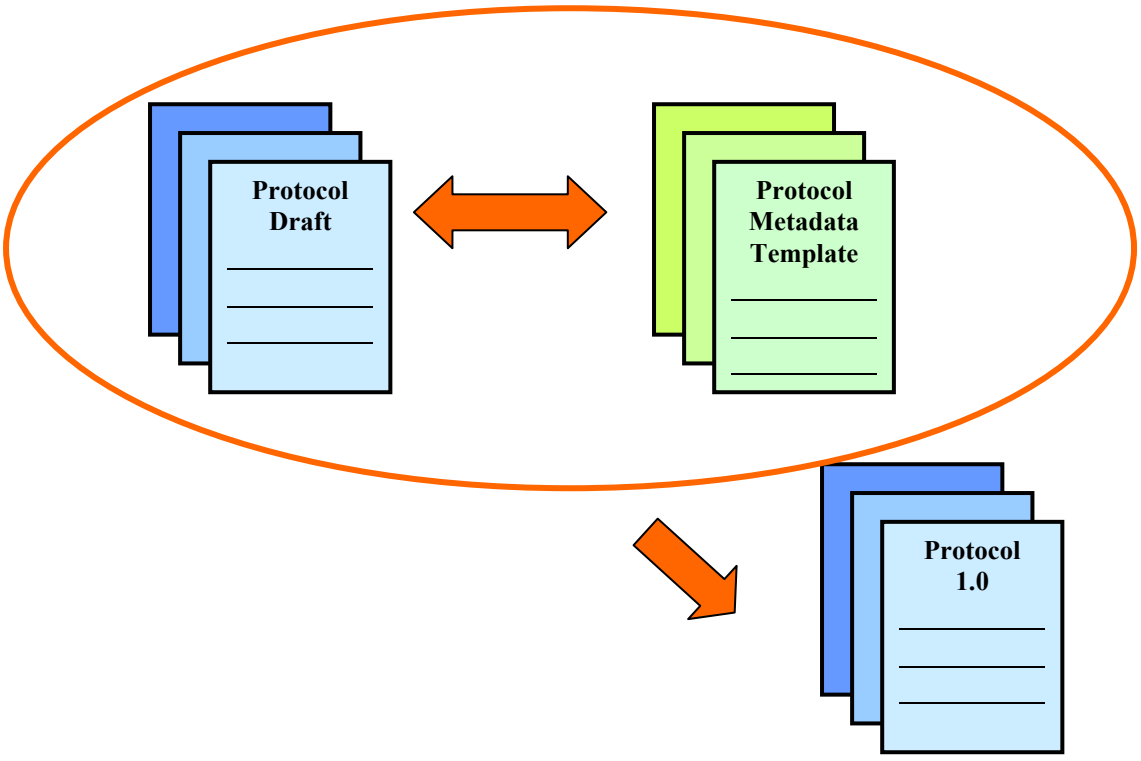



FIGURE 4

There was therefore a need to modify the process to ensure that the protocol standards were implemented correctly while the trial protocol was developed. First and foremost, this required management attention to enforce these processes. The process of writing a protocol was therefore changed and the protocol metadata template was incorporated into the new process. This process was reinforced by training, SOPs updates and KPIs. It should be mentioned as well that updating the protocol development process was not solely for the PMT, but it was also to ensure that entire process is more efficient and to reduce the number of protocol amendments.

**CONCLUSION**

Going forward, it will probably be difficult to further develop the protocol metadata templates, mainly because the combination of programming and an extensive code lists are testing the limits of Excel. Therefore, if the tool should be further optimised, then it would be essential to find a different program. Another reality is that this tool does require extra work that could be avoided. At the moment we enter metadata into several different databases – including the PMT, IMPACT, CDW. In order to work more efficiently it would make sense to enter trial metadata into one central repository that would then transfer the metadata to all the relevant programs. This would allow trial managers, data managers and statisticians to work more efficiently and ensure that data standards are used consistently across all programs and thus throughout a clinical trial.

Therefore in terms of improving the use of standards in trial protocols it is still necessary to optimise the current standards, technology, and processes. In terms of standards, it would be beneficial to include additional therapeutic areas. Furthermore, if the technology is improved, it should be done in a way that would prevent rework and support more efficient processes.

However, despite all of the possibilities for improvement, it is still important to remember that Excel has served as a good program to use to gradually implement protocol metadata standards. Since it was a low cost solution and easy to upgrade, there was more flexibility to optimise the standards, technology and processes. At this point, the use of metadata standards has been incorporated into the protocol development process and most protocol originators have been trained in CDW, metadata standards and the benefits. The greatest benefit to take from this is that a foundation for using metadata standards has been established in the protocol development phase. Going forward this foundation can be used to build a more complex system for implementing protocol metadata standards.

## RECOMMENDED READING

AD07 Metadata and Standard Programs
Marianne Carames, Novo Nordisk & Martin Lindhard, Novo Nordisk (Phuse, 2009, Basel)

TS06: Managing your metadata efficiently
Kirsten Langendorf, Novo Nordisk & Mikkel Traun, Novo Nordisk (Phuse, 2009, Basel)

AD03: Data Standardisation, Clinical Data Warehouse and SAS® Standard Programs
Jean-Marc Ferran, Novo Nordisk A/S, Mikkel Traun, Functional Architect,
Pia Hjulskov Kristenses, Business Implementation Specialist (Phuse, 2008, Manchester)

## CONTACT INFORMATION (HEADER 1)

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

Annabel Acs / Pia Hjulskov Kristensen
Novo Nordisk A/S
Vandtårnsvej 114
Søborg 2860
Denmark
Work Phone: +45 3079 6027 / +45 3079 9939
Email: aacs@novonordisk.com / phjk@novonordisk.com