

WHITEPAPER

Accelerated implementation of a clinical data warehouse

Table of contents

Executive summary	3
Introduction: Why a clinical data warehouse?	3
Clinical data warehouse value proposition	4
Data warehouse building blocks	5
Clinical data warehouse implementation	7
NNIT's approach to clinical data warehouses	8
Clinical data warehouse accelerators	8
NNIT's clinical data warehouse value proposition	10

Executive summary

This whitepaper explains the enormous benefits to be gained from a clinical data warehouse (CDW) and emphasises the importance of accelerating its implementation.

CDW is the natural solution to a rapidly growing need for organising, storing and sharing clinical data from different systems. The availability of clinical data across a company saves time, reduces costs and facilitates compliance. However, the decision to adopt CDW cannot be made without a clear idea of how to implement it. Implementation is an immense undertaking and should not be underestimated. If the CDW is to be delivered successfully and on time, it requires an implementation partner with a clear methodology, pre-designed solutions and a proven track record of implementation projects.

On the basis of extensive experience with clinical data management, statistical analysis and regulatory submission, NNIT has developed a comprehensive best practice CDW framework. This CDW framework differentiates itself by its clear focus on the acceleration of the implementation process and its unique capacity in metadata management. NNIT's approach to CDW substantially accelerates implementation and reduces the time spent on design, construction and testing by up to 50%.

This whitepaper details NNIT's approach and the building blocks of its CDW solution with a particular focus on the CDW accelerators.

Introduction: Why a clinical data warehouse?

Clinical data warehouse (CDW) is the clear response to a number of challenges faced by the life sciences industry. First, there is an almost explosive growth in the data resulting from clinical trials: more trials are conducted, each trial collects more data and trial designs have become more complex. Second, both companies and regulatory agencies are moving towards highly standardised study reporting and data exchange such as SDTM, and this is becoming more challenging as data volumes grow. Third, there is a definite need for adaptive business models when traditional life sciences companies outsource much of the clinical operation and statistical analysis, and data needs to be exchanged in a quick, structured and uniform manner. Finally, many companies take over new products or late stage development projects that then have to be transferred quickly into the company's systems in order for the data to be analysed and explored and to secure the continuation of the clinical development plan.

Since there is an increased pressure to get products to market, the above challenges result in a situation where more data needs to be handled in a shorter time frame. This increases the pressure on the data management and statistical departments. These departments regularly struggle to keep up with their daily work and often become bottlenecks in the release of the study data for any further usage. Consequently, valuable information may be unavailable and the life sciences company loses time, money and opportunities – and may even face regulatory interference.

Over the past decade, the industry has spent fortunes on systems that focus on improving the collection and handling of clinical data. However, inefficiencies in the overall clinical development process still need to be addressed. Many of the major life sciences companies have engaged in eClinical projects. These examine the entire clinical development process and develop road maps to introduce major changes according to how much the receiving organisation can absorb. Although most pharmaceutical and biotech companies have already implemented the bulk of the applications in the eClinical landscape, they still fail to realise the full benefits of the systems. The systems remain islands of data and are only available to highly specialised personnel.

The solution to these challenges is therefore the implementation of a CDW where data from different systems are organised, stored and made available to the entire company.

The implementation of a CDW is, however, an enormous undertaking, which should not be underestimated. It involves a number of key resources and places an even greater pressure on the organisation during the implementation project. One way of dealing with this increased pressure is to temporarily outsource parts of the data management and statistical analysis work to CRO's in order to free up the needed resources.

Another way of dealing with the overwhelming work load is to work with an IT partner with the right know-how and experience to deliver an already proven solution that can be adapted and scaled to the individual life sciences company. Focusing on methodology and experience can accelerate the implementation process to such an extent that it may not even be necessary to outsource some of the work to CRO's.

As described below, NNIT has the experience and methodology to ensure that the implementation of a CDW is not only successful, but also takes place in a highly accelerated fashion. Bearing in mind the acute time restraints present in the life sciences industry, this acceleration can only result in reduced costs for the life sciences company.

CDW value proposition

CDW's central benefit is that everyone can access the same data. When statisticians, statistical programmers, bio-modellers and pharmacokineticians have access to a statistical computing environment, including a standard program library, they can focus on data analysis, as it should be, and not on programming. When people involved in medical writing, pharmacovigilance and data management have access to a patient browser, they can focus on information analysis, as it should be, and not on discussions about data interpretations arising from discrepancies and lack of standards.

When medical writers, trial managers, health economists and epidemiologists can perform ad hoc analysis across trials, they can focus on future trial design, cost savings and adaptation of running trials instead of blindly repeating past mistakes.



Figure 1: Cross-organisational focus

A CDW project is therefore first and foremost a drive towards standardisation that enables:

- Better use of resources
- Standard exchange of data with CRO, partners, and regulatory agencies
- Cross-trial analysis and leveraged use of historic data
- Reduction in critical time path for statistical analysis
- Globalisation and knowledge sharing
- Compliance with regulations

Any life sciences company considering a CDW solution should start by analysing their current situation to determine which of the objectives listed above have the highest priority. This will provide a valuable focus for the implementation of CDW since it is rarely feasible to cover all objectives at the same time.

Data warehouse building blocks

A 'basic' CDW covers data load programs for most common data sources e.g. CDMS, EDC, SDTM/ODM, IVRS, Safety and CTMS, as well as transformation and enrichment of the data into a single standardised data model. The data load programs must include a load of the study metadata so that, at a later stage, it is possible to see features such as code-list and dictionaries version and trial design including trial arms and visit schedules. The CDW also includes a number of 'data marts', special collections of data organised for a specific purpose, such as a SDTM-data mart that can be exported or reported.

In addition, the CDW should always include user administration and access control. If data are loaded before unblinding, the system must be able to handle unblinding of data and access control to unblinded data.

As long as changes are not made in trial design components, e.g. trial CRF pages, code-lists and dictionaries, the solution works optimally. However, as soon as it becomes necessary to load data from other sources, add new therapeutic areas or change current standards, it becomes necessary to make changes to existing load programs as well as other programs such as SAS analytical programs and reports. It also becomes necessary to consider how to reflect these changes in the current data in the CDW.

Most likely, it will be necessary to have more than the original 'view' of the data, meaning the data set as it looked at the time of its collection. It might also be necessary to have a view that reflects the situation at the time of reporting or submission and perhaps also a third view for exploratory purposes. These views could be necessary because of factors such as changes in code-lists and dictionaries.

Faced with this type of requirement, it is necessary to investigate how to expand the functionality of the 'basic' CDW in order to meet all these needs without employing a legion of data managers and statisticians to maintain code-lists and programs. At this point, it becomes interesting to use metadata to drive data source mapping and clinical study reporting and metadata management to facilitate programming and study design.

Without the dynamic ability to shift study and trial metadata during the trial life-cycle, there is an enormous risk that the data grow stale, thereby reducing the data warehouse to a storage facility with little value to the users and to the company.

The metadata management application that needs to be added to the 'basic' CDW consists of three modules: a study metadata module, a source data mapping module and an administration module.

The study metadata module (SMM) maintains the following metadata: clinical metadata, study metadata, study design, visit structure, study flowchart, clinical metadata versions and cross-study metadata.

The degree of functionality that the source data mapping (SDM) module should cover depends on the number of data sources and the level of flexibility and variation in these data sources.

If, for instance, only data in SDTM version 3.1.1 are sourced, the possibilities are limited almost to the point where an SDM is not needed at all. If, however, data are sourced from Oracle Clinical and/or another CDMS, there will be variations in how the trials are defined and structured. In this case one should consider making the source data mapping dynamic and letting the SDM handle any necessary data conversion.

The amount of effort and functionality added to the SDM also depends on the strategy of migrating legacy data. It is commonly acknowledged that the migration of all historic study data into the CDW requires immense efforts. In some cases the data migration has been estimated to three times the implementation of the CDW itself. A smartly designed SDM will reduce this migration effort. The cost and benefit of the SDM should therefore be held up against the cost of migrating legacy data and the benefit of having legacy data available in the CDW.

Besides acting as a repository for study design and study metadata, the SMM also plays a pivotal role in the drive towards standardisation.

In order to maintain a high level of standardisation and actively pursue frontloading of resources, maintenance of metadata and the preparation of metadata for new studies should be done as early as possible in the trial design process. Preferably all new study protocols should be based on the metadata library and any changes necessary to accommodate new trial designs in the metadata library should be made and approved together with the internal approval of the study protocol. This process ensures that, first, all activities that can be front-loaded will be done and, second, that the CDW is ready to load study data as soon as data are returned from the clinic.

Once the SMM is available, it is possible to re-analyse the current processes in order to further leverage the system.

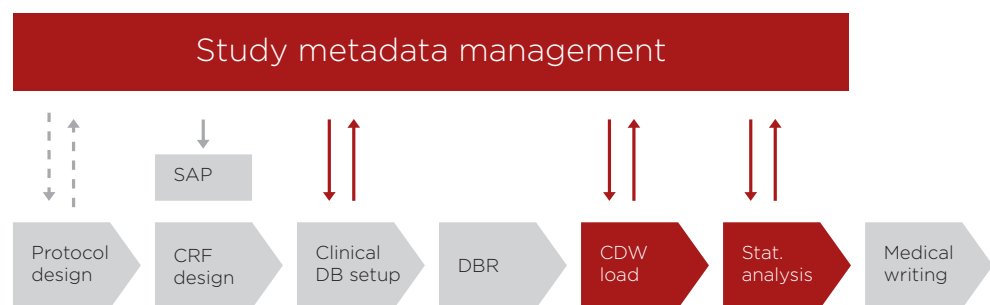


Figure 2: Process and data standardisation

The above process overview makes it clear that the study protocol's design should interact with the metadata repository. Some companies are investigating the possibility of implementing electronic protocol generators (ePG). This already makes good sense when it comes to standardising protocol design, but it makes even better sense if the ePG pulls its protocol components from the SMM as this will guarantee consistency between the way data is collected and the way data is stored and reported.

In this way, the SMM becomes the company's real repository for clinical trial handling and reporting.

CDW implementation

The objective of the CDW is typically to become the source of the statistical analysis and the foundation on which to create the analysis data set that is part of the submission. As a result the CDW has a potential impact on the final product and should be rated critical in terms of GxP. Consequently, the system has to be implemented in compliance with GAMP, ICH and 21 CFR part 11.

The typical implementation process for a CDW covers the following phases and activities:

Phase I: Definition of scope and business case, RFP and vendor selection

Phase II: Proof of concept

Phase III: Design, construction, test and validation, implementation and roll-out

As there are very few standard components in currently available CDW solutions, it is to be expected that some custom development is needed, especially regarding data load and metadata management solutions.

NNIT's approach to CDWs

In the past decade, NNIT has implemented solutions that support clinical data management and statistical analysis, and regulatory submission within the life sciences industry. Our involvement in such projects has given us extensive insight into the business processes that are critical to our life sciences customers. Since our customers are typically subject to similar regulation, we have discovered a convergence across our various implementation projects. Recently we have been the implementation partner on two successful CDW implementations in which we have combined our experiences and competencies to formulate a best practice CDW framework. Most noticeably, our CDW framework consists of building blocks that can be used to accelerate the implementation of a CDW.

CDW accelerators

The NNIT framework includes two parts: The 'basic' CDW and the metadata management.

The 'basic' CDW framework

Based on industry standards such as SDTM and best practice data models such as Janus, NNIT has designed an enhanced, normalised data model that allows cross-trial analysis and SDTM and ADAMs data set to be created. It is our experience and expectation that this data model is similar, if not identical, to what is designed across the majority of all current CDW projects. The data model is easily adapted if customer-specific needs are not covered. The NNIT framework includes technical design specifications, ER diagrams, installation scripts and test scripts, and test protocol templates.

The second accelerator in the CDW framework is a set of programs that read SDTM.xpt and Define.xml files and load the data into the CDW database. The programs are based on SDTM version 3.1.1 and use define.xml and the code-list/value level metadata contained in the metadata repository to verify that the data set can be loaded correctly and comply with approved company standards.

The SDTM load programs handle new or proprietary domains by storing all events in one EVENTS table. All findings are stored in one FINDINGS table and all interventions are stored in one INTERV table. The load program stores all additional supplemental qualifiers in a single SUPQUAL table.

The define.xml metadata are also loaded into the data model in order to correctly store historic values such as dictionary versions as well as trial design definitions.

The load programs include a set of logs where potential issues are kept for easy examination. In addition, the programs ensure that any data inconsistency is identified and logged. This situation could be caused by factors such as unscheduled visits or a broken trial arm. The NNIT framework includes technical design specifications, installation scripts, code-review reports, unit and integrations test documentation and test scripts.

The third accelerator is a database view that creates an SDTM data mart. This accelerator is highly standardised and includes technical design specifications, installation scripts, code-review reports, unit and integrations test documentation and test scripts.

The NNIT CDW framework is optimised to be integrated with Oracle Life Sciences Data Hub (LSH) and by implementing Oracle LSH further advances can be achieved. Together with Oracle LSH, the NNIT framework includes a number of LSH applications that produce a code-list report, generation of SDTM.xpt files and a simple set of reports in which demography as well as clinical events can be examined across trials. In addition, Oracle LSH offers a system optimised for handling clinical study data, which substantially reduces the effort spent on defining and building the administration and company structure needed for managing the CDW.

By using the NNIT standard procedure for installation, configuration and operation of standard software such as Oracle LSH, the NNIT framework covers guides for the full implementation of LSH including user roles, security system, organisational structure and validation guidelines.

The metadata framework

The primary accelerator is the SMM. This includes a metadata repository that is used during data load and other program development and execution.

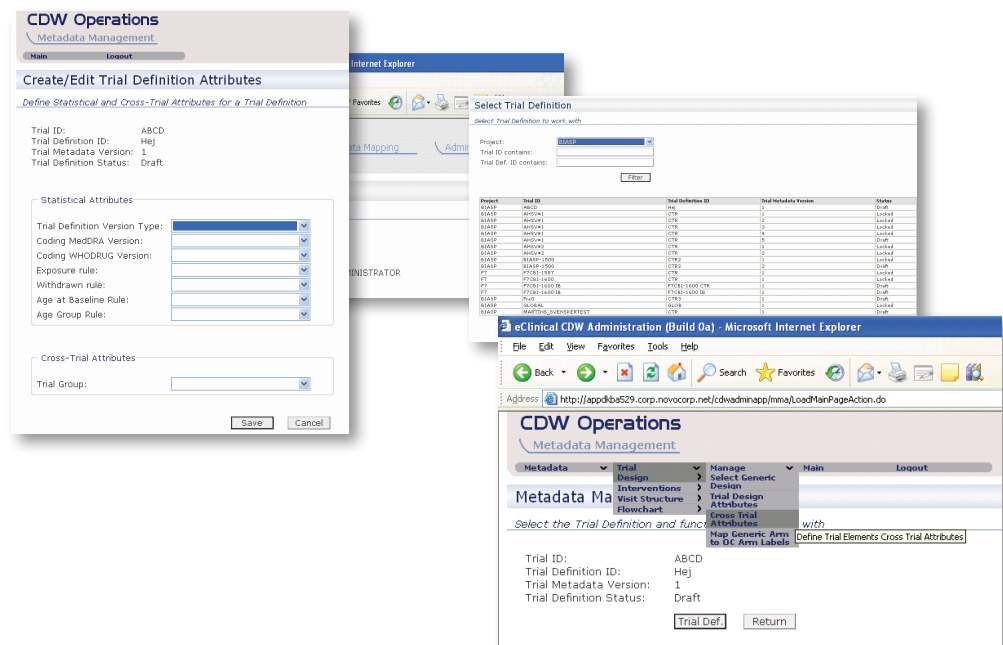


Figure 3: Metadata management

The SMM framework needs to be adapted to the customer-specific data model and loading programs. Consequently, this accelerator needs some additional technical specifications, design and code in order to be implemented. The NNIT framework includes technical design specifications and ER diagrams.

In addition to the SMM, the NNIT framework also includes a metadata administration module that maintains users and the security system. If the SMM is used without LSH, the administration module also includes a centre for managing load processes.

NNIT's CDW value proposition

With the NNIT CDW framework, it is possible to reduce the effort and work hours spent on the design, construction and testing by up to 45%.

Based on our two recent CDW implementation projects, NNIT has reviewed all design material with the purpose of determining the extent to which each component is customer-specific or if it is a general-purpose component that would fit any CDW solution. We have determined that approximately 85% of our 'basic' CDW design is standard and can be used by all pharmaceutical and biotech companies. We have also estimated that approximately 45% of our 'study metadata management' design is standard. The 'source data mapping' module is a highly customer-specific module as it is designed to work with a specific set of source data systems. Therefore, only little can be reapplied besides the huge knowledge that has been invested in this module.

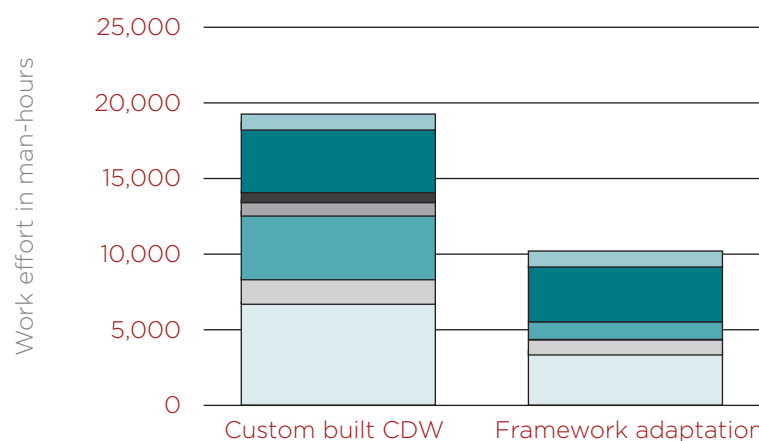


Figure 4: Project acceleration factor. Work effort in man-hours

In addition to the reduction in specification, design, construction and test of these highly standardised modules, NNIT can also reduce the time and resources spent on system validation activities. The NNIT framework includes a complete set of installation scripts and test protocols that are easily adapted and executed.

Finally, a framework like the NNIT CDW framework reduces the risks of project budget overruns, sliding milestones and faulty design: more than two thirds of the solution is built on a proven design by a partner with a proven track record in the implementation of CDW solutions in the life sciences industry.

NNIT is one of Europe's leading consultancies in the development, implementation, validation and operation of IT for the life sciences industry. We create value for our clients by treating their IT as if it was our own, and of course, we meet the industry's strictest requirements for quality. For over a decade, we have applied the latest advances in technology to make our clients' software, business processes and communication more effective. NNIT employs nearly 1,300 people and in 2008, our turnover exceeded €185 million.

www.nnit.com