PhUSE 2010

Paper DH11

DATA WAREHOUSING FOR THE REPORTING AND MANAGEMENT OF CLINICAL DATA

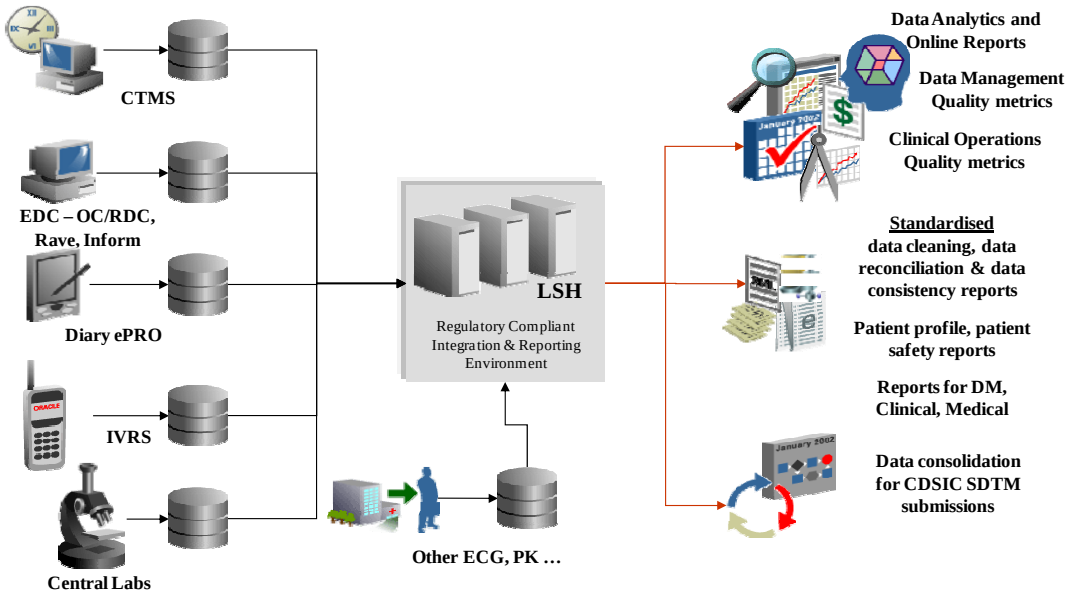
Robert Ellison, ICON Clinical Research, Dublin, Ireland

ABSTRACT (DATA WAREHOUSING FOR THE REPORTING AND MANAGEMENT OF CLINICAL DATA)

The purpose of this paper is to share ICON's reasons for, and experience of, planning and implementing a data warehousing solution

INTRODUCTION (DATA WAREHOUSING FOR THE REPORTING AND MANAGEMENT OF CLINICAL DATA)

The primary reason for ICON deciding to invest time in implementing a clinical data warehouse infrastructure was to enhance the analysis and delivery of clinical trial data to sponsors. The decision was made in the beginning of 2009 to use Oracle's Life Science Data Hub (LSH) to deliver a technology foundation to support clinical data integration and reporting within ICON. The high level goal was to have one central location in which data from all sources (internal and external to ICON) could be stored, using a global approach to processing and analysing the data.



Over the past 12 months a team of IT and Data Management staff have installed the warehouse structure, designed global standard data structures and program code, arranged global library and study specific directory structures and created/updated the processes involved in setting up and maintaining study data within this environment.

Phase I of this initiative went live on 30th July 2010. Data Management's programming and study team staff are now able to simultaneously access clinical trial data from anywhere in the world and provide a more globally integrated data management solution to meet sponsor needs. Future phases will focus on extending the user base across the larger internal organisation and out to sponsors to ensure consistent and cohesive management of clinical data.

IMPLEMENTING A WORKABLE DATA REPOSITORY SOLUTION

The effort required to put a working data warehouse in place required focus on two main areas. From an IT perspective the appropriate hardware and software had to be in place to facilitate the Life Sciences Hub for production, test and development environments. From a data management perspective processes for loading data into the warehouse environment, transforming it and producing output had to be developed and validated.

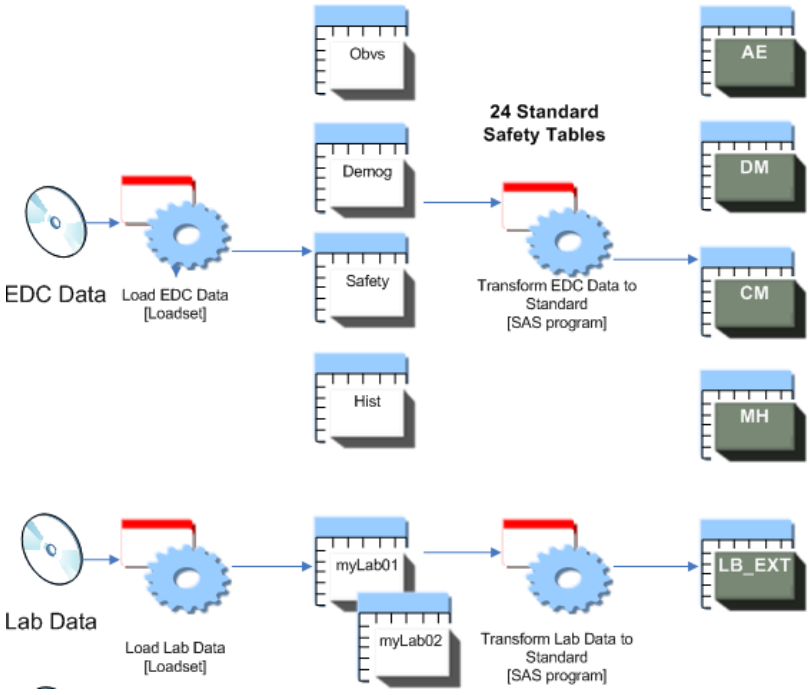
HARDWARE & SOFTWARE CONSIDERATIONS

Taking into consideration that the LSH Production Environment must be available 24x7, the architecture design includes Multiple Application Servers and Oracle Real Application Cluster providing a high availability solution, built on proven Oracle technologies. The Test Environment is a similar architecture to the production environment to ensure that any testing performed in this environment will closely match the Production environment. The sizing for the test environment was significantly lower than for production. The development environment is an area for ICON programmers and developers to test system functionality and program code without impacting on the production and test environments. The sizing for the development environment was lower than the test environment taking into consideration its usage.

DATA, PROGRAMMING AND STRUCTURAL STANDARDS

The first consideration from a Data Management operational perspective was how deal with the different data sources that we were loading into the LSH environment. ICON has three core Clinical Database Management Systems to take into consideration, Oracle Clinical (paper based and RDC studies), Rave and InForm and a wider range of supplemental sources such as Lab, IVRS, ECG, PK, Diary data. Structures are in place to house this raw data in standard LSH organization hierarchies so that they could be easily identified for the next step – transformation into standard data structures. The raw data is categorized by type with one study specific workarea for data coming from the CDMS, a workarea for data provided by laboratories, IVRS and so on.

Standard data tables were put in place based on the Study Data Tabulation Model with view to these housing data from all raw data sources. The structure and versions of the tables are stored in a global library and can be copied down to a study specific location before being populated. To transform the data into the standard structures a SAS program template was devised to aid programmers best fit the data in as an efficient manner as possible. This provided in interesting challenge which are covered in the CRO – Specific Challenges section of this paper. Due to the CDMS system and study specific nature of the raw data the transform program requires a certain amount of manual programming to ensure the data is mapped correctly to the standard structures, which is way the code is stored in the library as a template rather than a complete and validated piece of programming code. This means that the transform step has to be validated on a study by study basis.



Data cleaning and reconciliation reports are also developed using SAS within the LSH environment. These reports are designed to take data from the standard structures and therefore do not require to be updated to meet study

specific requirements. The code to generate these outputs is validated once and then moved in to the global library. Further validation is not required once the program code remains the same. There is still a requirement from time to time to produce study specific outputs (and there probably always will be) and this can be facilitated within the LSH environment. Study specific code can be written to reference the raw data loaded in before the transform step is implemented, the transformed data or a combination of both.

All processes within this environment must pass through the system life cycle, going from a state of development during the set up stages, to QC during the validation process and finally into production. While this is a process that we follow outside the confines of a data warehouse there is a significant advantage to having each step monitored and controlled within a single environment. As the program moves through the lifecycle it's status is recorded in the audit trail.

SHORT AND LONG TERM BENEFITS OF HAVING ONE CENTRAL LOCATION FOR THE STORING OF DATA

One of the major benefits for Data Management in using one central location for storing data is the reduction in time required to agree on and set up study data structures, programs and processes. A standard directory structure for the data is copied down from a global library along with standard data structures and program code to generate data cleaning and reconciliation outputs. This will significantly reduce the amount of time required to put these processes in place at a study level.

The standardisation of the raw clinical data can also be viewed as step to simplify the data. By mapping system/study specific data into a standard structure we are creating a view on the data which will become very familiar to the teams in Data Management, Biostatistics, Clinical Operations, etc. who work with it on a regular basis.

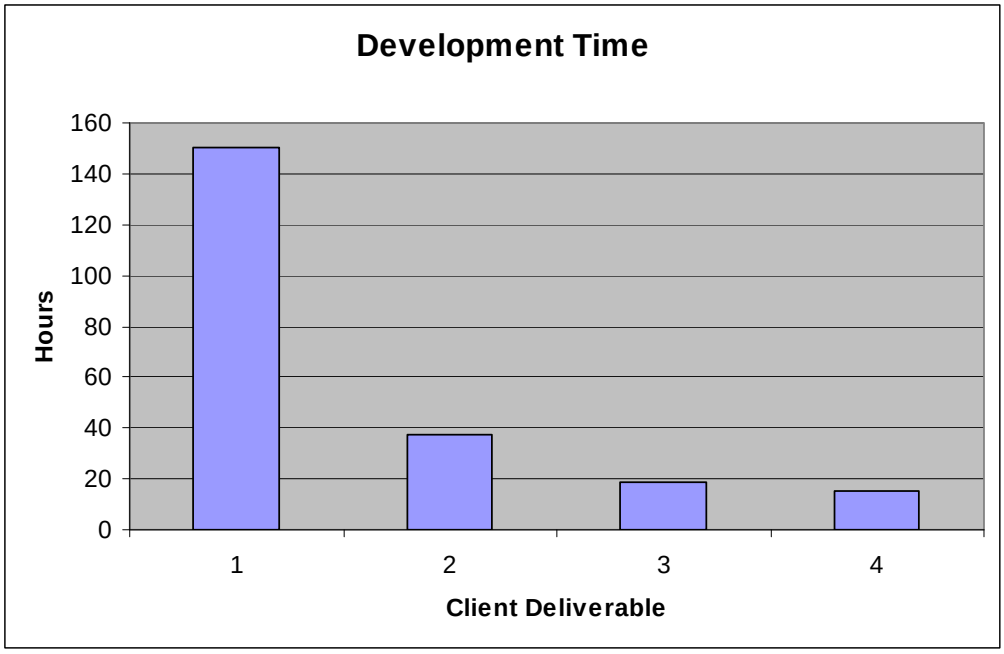
Phase I of the Data Hub roll out in ICON has been focused on actual clinical data and there are real benefits in the short term relating to the standardisation of operational reporting. There are a large number of data cleaning and reconciliation reports which can be reused on study after study without any further programming or validation effort. In the longer term as we turn our attention to metric information there will be other benefits in terms of producing standard management reports. Reports relating to study status and performance will be available at subject, site, study, program and sponsor level. In the we faced difficulties in reporting on data from different systems and sources we will be able to report freely at multiple levels.

To fully utilise the ability to report at multiple levels we are in the process of developing a clinical data reporting schema which will allow us to use more sophisticated business intelligence (BI) reports. The reports created to date will allow users to view information in graphical or tabular form and drill down through a sponsor, program, study, country, site, subject hierarchy to get to the required level of detail. Looking to the longer term metric information is ideally suited to BI reporting and that is a future area of focus. The Clinical Data Analytics tool developed by Oracle already provides us with intelligent reporting on metric data and is something we are looking to develop further.

As a global company the ability to have one global repository from which all regions and departments can work from will give us a greater ability to manage trails across data centres, scale people, processes and technology.

CRO SPECIFIC CHALLENGES

The need to standardize data across multiple systems and data structures has been the biggest single challenge in centralizing our clinical data. Working for multiple sponsors and receiving data in from multiple sources means that we are limited in what we can do to standardize data at source. When source data does not adhere to a standard format the actions required to process this data will have to be study specific to a certain degree. A lot of consideration was given to the amount of time required to standardize the data at a study level in comparison to the time saved by removing the need to program and validate study specific code to produce outputs for reporting and data cleaning purposes. Case studies were carried out to assess the efficiencies of using standardized data structures and from these test it became apparent that we would realize real benefits over a reasonably short period of time. Below is an example of the amount of hours required to program the same set of reports for four different clients, the first deliverable including the development time required to get the standard code and data structures in place:



FUTURE PLANS

As the volumes of data loaded into the data hub increase we fully expect to see the need to review our structures and standardisation process. This will be an ongoing process. In addition to maintaining what is already in place there are plans to further extend the use of LSH within ICON. There are several other department areas managing data, programming against data and delivering outputs to sponsors, which will be potential candidates for LSH in the future. We plan to expand on the range of BI reports available , patient safety reporting and patient profiling being examples of the areas we are looking into. There will also be a focus on overall discrepancy management solution to see if that is something that would be beneficial to have inside the LSH environment.

.

CONTACT INFORMATION

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

Robert Ellison
ICON Clinical Research
South County Business Park
Dublin 18
Work Phone: 00353 (01) 2912405
Email: robert.ellison@iconplc.com
Web: www.iconplc.com

ICON
A Symbol of Excellence

Clinical Research