PhUSE 2012

Paper DH05

Improving Data Management with Systems Integration and Next Generation Data Flow

Richard Davies, Oracle, Reading, UK

ABSTRACT

EDC systems and other automation systems have become ubiquitous, creating an inherent need to connect multiple systems in order to maximize investments in technology. Managing the data flow between these systems may be daunting to both new and experienced study designers; and determining the correct questions to ask when designing a highly integrated study is not intuitive. This presentation will identify key decisions necessary when planning a study that will integrate IVRS data, EDC data and ePRO data and compare benefits / risks of designing data flow between systems. Additionally, the technological innovations make it possible to move electronic data to the cloud, bringing new possibilities, efficiencies and scalability. With the new possibilities come new challenges, namely: how do we keep the data safe, secure, and protect the privacy of patients in this brave new world. The speaker will discuss considerations and next steps to prepare data for the cloud.

INTRODUCTION

There are several challenges we continue to face as we execute our clinical trials. One is be increased diversity of data sources associated with anyone clinical trial. Another is our ability to choose diverse technology to support the execution of a clinical trial, especially through the use of cloud based solutions that can be consumed on a trial by trial basis. Add to those the potential to outsource components of a clinical trial and internally these factors contrive to complicate our lives through complex integration and data reconciliation challenges. This paper discusses these challenges and how technology can help or hinder us in these situations, as well as discuss some of the solutions that may be available to us.

DATA SOURCES AND OUR TECHNOLOGY FOOTPRINT

The complexity of the clinical trials we run has increased over time as we become more sophisticated in the data we can collect for analysis and technology providers have innovated to respond to our needs. 15 years ago the technology footprint associated with anyone clinical trial extended to a Clinical Data Management System and Statistical Analysis software. Today we may use technology to control randomization, drug supply, transport laboratory data and collect data from sites and subjects independently, amongst others. There are many choices and questions that need to be answered when we look at implementing these diverse technologies on any one clinical trial. We may choose to minimize the technical complexity by avoiding integrations and dealing with all the diverse data sources internally as the data comes back in house. This approach reduces the complexity of study startup and accelerates the time it takes to bring a study life. Alternatively we may choose a highly integrated approach where more effort is invested up front to increase our integration and minimize our downstream work with the reverse scenario whereby we typically complicate and extend the study setup phase. There are pros and cons with each approach and we'll examine some of the influencing factors that drive the choices we make one.

A PRACTICAL EXAMPLE

Let's take an example of a clinical trial that will run using Electronic Data Capture and Interactive Voice/Web Response Technology. Perhaps the trial is in Oncology and so we know we'll have a relatively high number of dropouts. In this case a study drug is very expensive to manufacture and owing to the characteristics the drug so is the placebo. Therefore not only do we want IVRS technology, we also need integrated drug supply management capability. In the use of these systems we can think of scenarios where the data in one system would, in an ideal world impact another. For example when a subject is randomized in IVRS then the drug kit and randomization numbers could transfer to EDC. If a subject becomes a protocol violation and is discontinued then associated drug supply should stop.

OPERATING IN ISOLATION

So what would be the pros and cons of running the EDC System and the IVRS System in isolation on our clinical trial ?. Let's consider the investigator/site view first. Operating the systems in isolation will complicate the site users lives because they will have two systems to access with two username and password combinations and potentially multiple users accessing the to the systems at the same time. There will be a degree of data duplication as visits

are recorded in the IVRS System and subsequently transcribed from paper into the EDC System. This introduces risk of inaccurate data being entered into the EDC system through transcription errors although this should be picked up by the Monitor when reconciling with the paper source during source data verification. From an internal perspective we will need to reconcile the data between the two systems to check that all the common data points agree. To be proactive about this and catch such issues as soon as possible requires both systems are able to provide us with data or extracts that can be processed internally. We have a reliance on timely internal resource and a inappropriate communication mechanism back to the site to resolve any problems that may occur. We also have to consider the naming conventions used for the variables in both systems especially if using technology from two different vendors. It's quite common for different naming conventions to be used and subjects could be 'Subject', 'Subj', 'SubjectID', 'usubjid' to take just one variable that might be common across systems. It unlikely that we will be able to take the data sets and reconcile them without some reprocessing or some upfront configuration work to enable that to happen.

INTEGRATION

To minimize some of these data cleaning challenges we could integrate the two technologies at a minimal level. This may be the only option if we have selected technology that is from two separate providers whereby the sophistication of the technology and integration capability has to be configured to the lowest common denominator, typically file based exchange of information. In such situations we are still using two separate systems with two separate interfaces that still complicate the site users lives by providing two separate logins requiring two sets of authentication credentials. This in turn means that both our applications can be used in isolation and we still run the risk of data inconsistencies finding their way into both systems. When we consider the usage of those systems it may be safe to assume that the IVRS System will always be used before the EDC System in terms of data collection. Therefore a minimal integration could involve the uni-directional flow of data points from the IVRS System into EDC. The data points collected within IVRS can be pushed into the EDC System including any relevant clinical data points and the randomization and visit information which can be used to pre-populate the eCRF pages and predefined visits as required. We also have examine the question of eCRF design and decide whether fields that transfer from within the IVRS System shall remain read only within the eCRF, or whether the site can be given data update privileges to these fields. If data update privileges exist then obviously we have the scenario whereby data can be modified and discrepancies exist between the two systems. If no change control is granted within the eCRF then we are enforcing the process whereby the IVR System has to be used prior to transcribing information into the eCRF system if information such as visit date, which is typically key item, will be transferred. Therefore we need to ensure that the integration between the two systems is timely and robust so as not to block the site user's interaction with the eCRF because data has not transferred from the IVRS System.

TECHNICAL APPROACH TO INTEGRATION

The sophistication of the integration technology between the systems can contribute to the effort involved in configuration. The most unsophisticated integration will use a series of triggered data exports, ftp exchanges and data imports to move data between one system and another. In a minimal integration approach above, our IVRS system will need to be configured so that it exports the relevant data points when IVRS transactions are complete. As the transaction complete this will trigger a file export in the format previously negotiated with the EDC System provider. Once the export has taken place and the file of data exists on the IVRS server, a scheduled job or sniffer process will then invoke to copy the file of data through a secure pipeline to the EDC System. Once received another scheduled job or sniffer process will detect of the new file of data received from IVRS and trigger a data import into the EDC System. The risks of this approach are that the scheduled jobs or sniffer processes stop working, the data export isn't in the correct format and so won't import into the receiving system or the physical link between the servers is down and data transfer cannot complete. Therefore we also need a notification and monitoring process involved with this type of integration and resource to check the output to ensure that integration is working and technical issues are not preventing data transfer. It is to the frustration of the site users when expected behavior between the two systems does not occur especially if they become blocked from working in one of the systems because data has not transferred successfully.

EXTENDED INTEGRATION

In the minimal integration approach described above we're enforcing a workflow on the site staff based on the underlying integration characteristics, i.e. Data only flows one way, so this dictates process. A more complete integration would see a bi-directional flow of information between the two systems that would enable the site users to work in isolation with the key data points transferring to the system where they do not currently exist. This may be better in a situation where visit information is typically stored in the IVRS System but he is not integral to the operation of that system. In this visit date scenario, having this information transcribed from a worksheet into the EDC System is perhaps the only required process needed at the site. With a bi-directional transfer of information this can be transferred to the IVR System saving the sites duplication of date entry. If a bi-directional transfer of information is required then the same risks are inherent if the technical approach described above is facilitating the data transfer. We also do have a think carefully about whether the same data points is editable in both systems so that the integration has the appropriate logic embedded to check whether data already exists in one of the systems when a data transfer takes place. We also have to decide whether we would allow data to be updated if it already

exists. With a technical solution described above we are reliant on such logic happening in batch and not in real time so that the end users are unaware that the usage of one or more of the systems is potentially causing an issue.

It's worth bearing in mind that in this particular example some of the data points that can be exchanged could have a bearing on randomization within a clinical trial. In these situations we have to be even more careful to properly control exchange of data points between systems and ensure integrations have the appropriate logic to prevent changes that may stimulate undesirable behavior in one or more of the systems.

HOW CAN TECHNOLOGY HELP

Technology developments have made it possible to deal with some of these challenges predominantly by giving us a mechanism to exchange data between systems in real time. The development of web services has given us the option to leave behind the file based transfers of old and implement data exchange that can be executed as soon as the transaction takes place in the instigating system. A web service is effectively a plug into an application that can be used securely over the Internet by any other application that knows how to use it. The advantage of web services are once they are defined they are maintained as standard even if the underlying application changes radically, so the operation of the web service stays the same and any other technology that knows how to use it can continue to do so. In our example our web service could help the integration in the following ways. Firstly because these the web services are real time transaction we get immediate feedback through the interface to the user who is acting on data. So for instance, if the user goes to the EDC System and tries to randomize a subject that is not currently randomized then the EDC System can reach out through the web service and initiate the randomization from within the IVRS System, storing the randomization results back in the eCRF in real time. Should another user been logged into the IVRS System and randomized the same subject at exactly the same moment they will simply be told that the subject is already randomized. So now we have no concerns over which system is used to trigger the activity because the integration has embedded logic and protection as the transactions occur. Web Services give us real time transfer of information within the initiating transaction, real-time reporting of information back to the end user and the least technically complex integration which is less prone to failure. Web service can also be used to exchange user credentials so that we can authenticate a user against multiple systems through a single login, hence removing the need for sites users to access multiple systems and remember multiple sets of credentials.

We also have a more secure mechanism to transfer or our data because the transfer occurs through a secure http connection directly from one application to the next. In the less technically complex scenario of file based transfer described above, source files and log files tend to remain on the respective servers in case the integration breaks down and we need to re-execute it. With these data files existing on servers accessible by internal vendor staff, we also have to consider the security and access levels to the servers so the blinding and sensitive information access control is preserved. No such issues exist with our web services approach.

BACKEND INTEGRATION

So far we've been considering how integration can help to control system behavior on the front end and lessen the reconciliation and data management overheads associated with exchanging data between systems. If we expand our scenario to include the use of electronic devices to record patient diaries or quality of life questionnaires then we introduce a possible integration that could happen on the back end. In this situation we need to ask ourselves whether there are any benefits of taking the diary or QOL information and making available to the site users. In some situations there is a desire for the investigators to see this type of information and that dictates an integration directly with the EDC System. However it may be that the integration requirements are not for the benefit of the site users and purely in the interest of reconciliation of the data sets from a data management perspective. In this situation a data warehouse or clinical repository solution which facilitates backend processing may be used as an integration solution to streamline the integration and subsequent processing of data from the systems. If it's possible to electronically deposit data from the EDC System and ePRO System directly into our clinical repository on a real-time or batch basis, then we can pre-configure the repository to handle the data reconciliation work and flagging of associated issues so that it functions as a lights-out operation. By configuring a clinical repository to act on new data as soon as it received, there repository can trigger a chain of events which implement our ETL routines, reconciliation routines and conversion to data standards as well as issue an appropriate log of all issues found along the way. Using this approach enables us to lessen the of the work involved in getting clinical trial technology 'live' in the first place by not implementing complex integrations on the front end. It allows for more simplistic back and integration coupled with automated processes for checking and dealing with the subsequent data that lesson the data management burden typically associated with reconciling data from multiple systems.

CONCLUSION

In conclusion we can summarize that there are multiple ways in which we can deal with clinical data arriving from different source systems. More sophisticated integrations on the front end will help prevent some of the issues that could potentially occur with clinical data but will complicate the set-up and go-live of a clinical trial. We have to think about naming conventions, data access control, site processes, supporting technology, log files and reports and factor these into our decision process as well as looking for scenarios where we can deal with data more appropriately on the back end, implementing the technology we have at hand to try and automate the process where possible. The more sophisticated and capable the integration then the less issues with the data we will have to deal with as a data manager. As these types of web service interfaces become more prevalent it will become easier to deal with data integrations between different systems. These integrations may not remove a need to reconcile data completely but they should lessen the situations where we find discrepancies between data from different systems, as well as providing usability enhancements for our user communities.

CONTACT INFORMATION

(In case a reader wants to get in touch with you, please put your contact information at the end of the paper.)
Your comments and questions are valued and encouraged. Contact the author at:

Richard Davies
Oracle, Health Sciences Global Business Unit
Oracle Parkway,
Thames Valley Park,
Reading, RG6 1RA, UK
Work Phone: +44 1189249732
Fax: +44 1189249732
Email: rich.davies@oracle.com
Web: www.oracle.com