

Remote Data Capture in Clinical Trials

An Oracle White Paper

Copyright© 2001 Oracle Corporation

Keith Howells
Senior Director, Oracle Pharmaceutical Applications
Version 1.3 - February 2001

Introduction

When you conduct a transaction with any kind of business these days, the information is entered directly into a computer system. Book an airline ticket, rent a car, buy a washing machine and the transaction is recorded real-time into a computer. The reasons for this are clear and compelling : the information can be checked for accuracy and completeness before the transaction is finalized, the transaction may immediately trigger other processing such as ordering and invoicing, and the information is immediately available for tracking and querying by the supplier and customer alike.

Yet during clinical research in the pharmaceutical industry, most information is initially recorded on paper and only transcribed to a computer system some time later. Some of the reasons behind this are logistical, because a patient's bedside is not the ideal place for a keyboard, and some are more related to the availability of appropriate technology. However, regardless of the reasons, the classic problems of paper-based data collection are magnified within the pharmaceutical industry. If the data is incomplete or inconsistent, it might be impossible to retrieve or correct the data when the problem is only identified some weeks later; all communication and data corrections must be scrupulously documented according to industry regulations; and busy medical professionals have to spend an inordinate amount of time on tiresome paperwork.

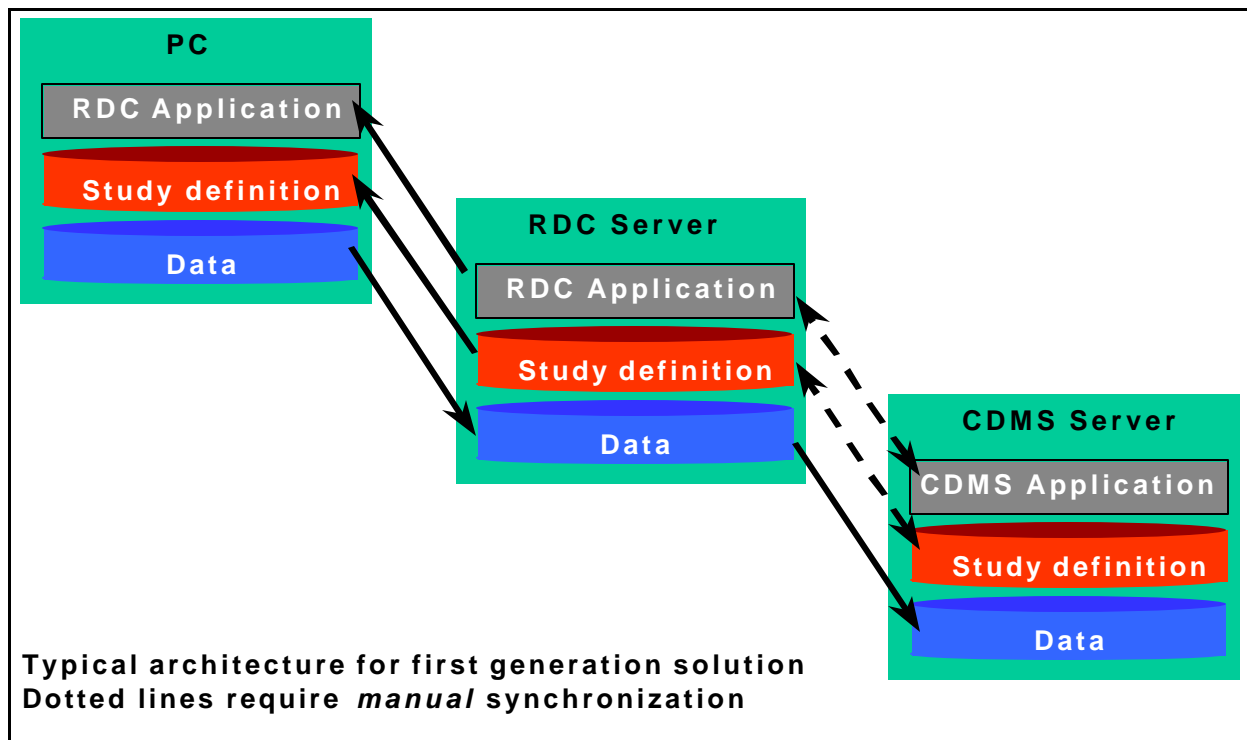
As the costs of clinical research continue to climb, and as more and more demands are made on the time of medical professionals, the paperwork logjam has emerged as a major area of attention. Something has to change.

The Failure of First Generation Systems

The impediments in paper-based data collection have been recognized for over a decade, and most companies have conducted a handful of pilot studies using some form of Remote Data Capture technology. The approach has typically been to pre-configure PC's with the Remote Data Capture software and the protocol definition, and ship those PC's to the investigative sites. The data is then entered locally and transmitted to a central database using a dialup connection. While this approach is outwardly appealing and the user interface for these PC-based systems has often been very attractive, a number of operational problems have emerged in practice.

The chief cause of these problems has been the dispersion of technology and the consequent loss of control when this technology needs to change. This approach requires a physical distribution of the RDC software, the study protocol and the actual patient data, any one of which may need to change on a frequent basis. Broadcasting a bug fix or a protocol amendment to hundreds of remote, disconnected PC's is not easy; synchronizing the data models and the edit checks between the PC system and the central database is not easy; and catching a data problem in the central database and coordinating a correction via the remote PC is not easy. At any one time, it may be unclear where the "gold" copy of the data resides and whether the central database is up to date. Resolving these issues requires a lot of attention from skilled information systems and data management personnel, which can outweigh the benefits of capturing the data electronically in the first place.

Periodically, the results of an RDC project would be presented at an industry conference. Almost always, the speaker would conclude that while aspects of the study conduct showed great promise, they had not been able to demonstrate any significant improvements in cost, time or quality. The attendees, relieved that they wouldn't have to do anything on this front for another year, would return to their desks and carry on shuffling paper.



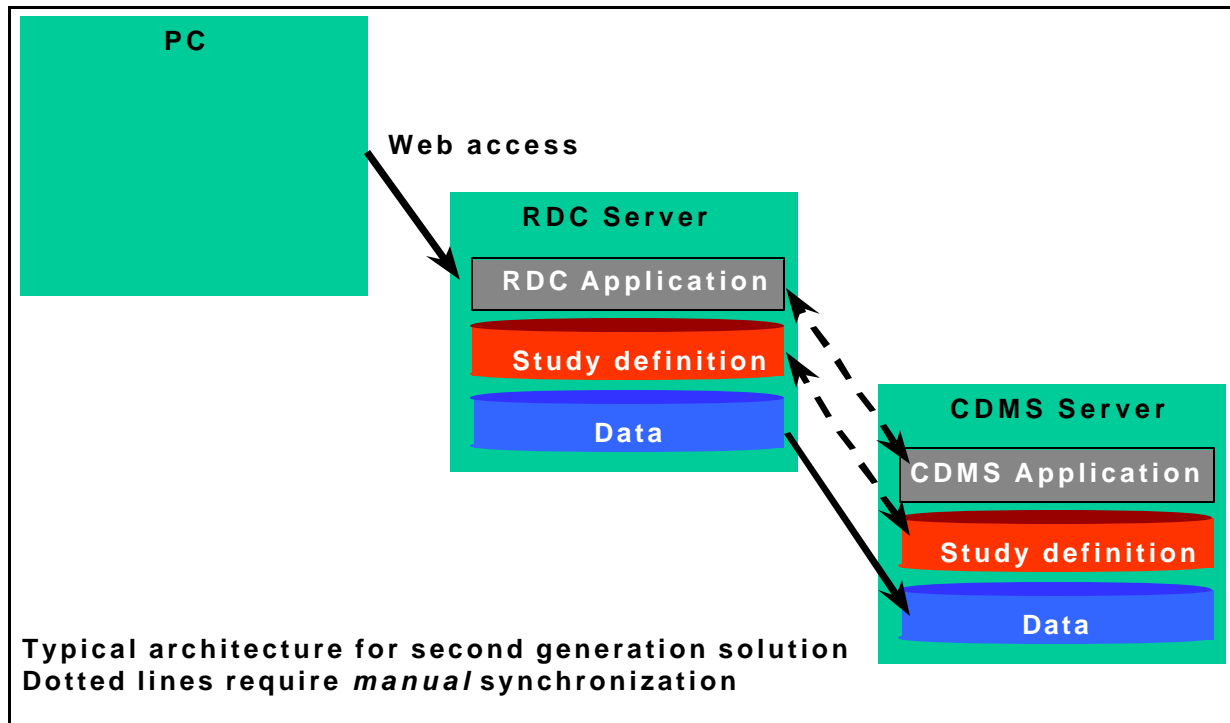
Second Generation Systems - Enter the Internet

In the late 1990's, a new approach to Remote Data Capture emerged by using the Internet. This approach immediately solves the problem of technology distribution, in that there is only one copy of the data collection software, only one copy of the study protocol and only one copy of the patient data. All that is needed at the investigative site is a standard PC with a standard web browser and a connection to the Internet. This is something that is becoming standard fare anyway for other applications such as insurance billing. Many companies have conducted at least pilot studies using the Internet over the past few years, and there is growing consensus that Internet-based systems work better than systems based on distributed PC's.

However, there are still a number of limitations with these early Internet solutions. Some edit checks may be too complex to program in the RDC system, or may be too resource-intensive to run in real time; some edit checks may require access to data which is collected in a different system, such as a central lab system; analysis of adverse events, medications or surgical procedures requires a full thesaurus management system, while analysis of lab data requires a lab reference range system. In short, to be fully useful, the Remote Data Capture system requires all the features of a classic clinical data management system. If any of these features are missing, then the data must be transmitted from the

Remote Data Capture system to the central data management system, and the problems of data dispersion and synchronization return.

Another issue is that these standalone Internet solutions imply an “all or nothing” approach. All sites in a protocol must use the Remote Data Capture system, since there is no way to process data collected on paper unless the study is to be independently set up in a separate system. This restricts the use of Remote Data Capture to technologically-advanced sites in technologically-advanced countries. Other studies must be done exclusively on paper.

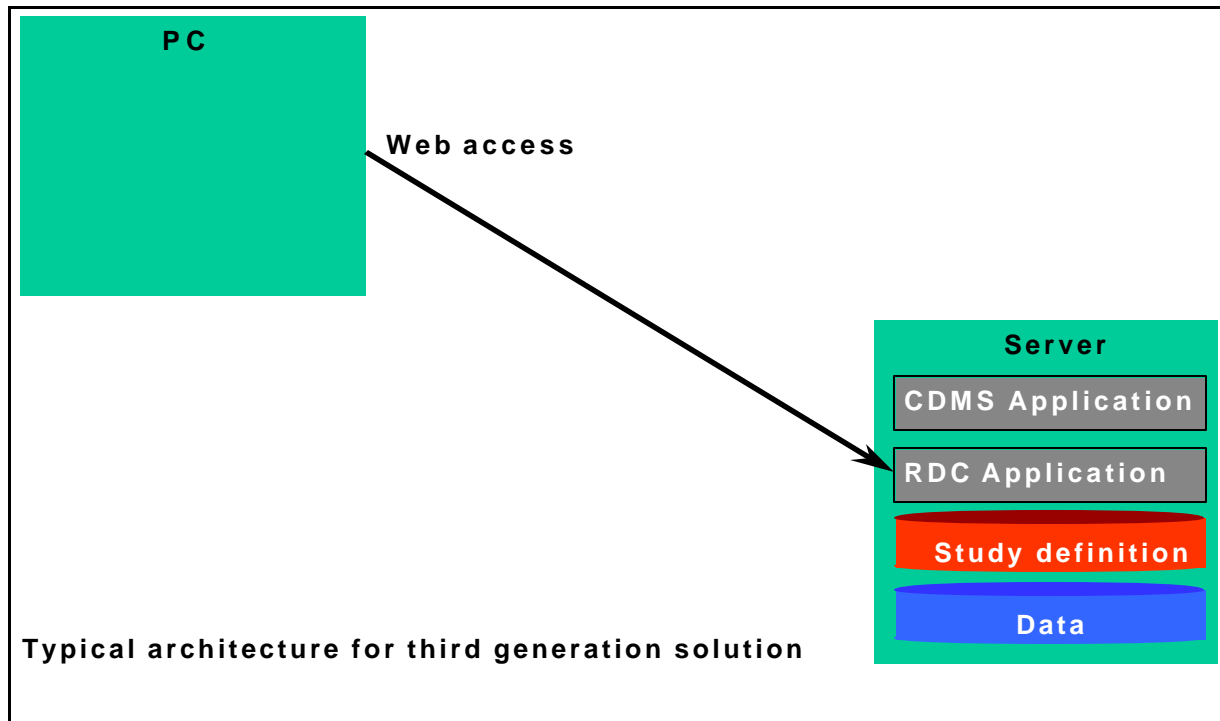


Third Generation Systems - Fully Integrated

Oracle’s approach is to run the Remote Data Capture system dynamically against the central clinical data management system - Oracle Clinical. This finally solves the problems with technology distribution and database synchronization, in that there is one copy of the study definition, one copy of the Remote Data Capture software and one copy of the patient data, all in a database that can handle both the front-end data capture and all the back-end processing. Oracle Clinical includes a powerful batch validation engine, a batch data loader, a randomization system, a thesaurus management system and a lab reference range system, with the results of these systems accessible to the Remote Data Capture user. For instance, central lab data could be batch loaded into Oracle Clinical, the reference range system could flag out-of-range values, and the investigator could use the Remote Data Capture interface to indicate which of the out-of-range values was clinically significant.

The integrated approach also means that a study is set up only once, regardless of the method of data collection. Standard tools for creating a study from standard components are available, and any subsequent protocol amendments are easily incorporated. An individual study may use a mixture of Remote Data Capture and paper-based collection, or data could be collected entirely on paper and the

Remote Data capture system could be used purely for query resolution. This “mix and match” capability gives the study sponsor great flexibility.



The User Interface Challenge

The idea of an integrated system for Remote Data Capture and clinical data management is appealing, but it will not work if the user interface is inadequate. Traditionally, data management systems have been fully-functioned but somewhat difficult to use, with heavy training requirements and a steep learning curve. This may be acceptable when the end-user is a full-time data management professional, but it is not acceptable when the end-user is a study nurse for whom computer systems are a sideline. Clearly the Remote Data Capture system must be an order of magnitude more friendly than the back-end data management system.

One approach is to make all the operations very procedural. For instance, buying products over the Internet follows a fairly simple pattern - choose products, give delivery information, give credit card information, confirm the order. Simple data entry could also follow a pattern - choose a patient, choose a visit, choose a CRF, enter data, and move to the next CRF. However, the full range of possible operations makes such procedural techniques difficult to implement. The system doesn't know if the user wants to enter new data, find existing data, search for open discrepancies or perform electronic approvals; indeed the user might not know what they want to do with a particular piece of data until they find it. Therefore, having a pre-set list of procedural steps in the software can actually become an impediment to productive work.

Oracle sees the user interface as being more akin to a word-processor. A word-processor simply wakes up and sits there with a flashing cursor, inviting the user to decide what to do next. Yet a word-processor is considered user friendly - indeed, attempts to guide the user through procedural steps using

the “dancing paperclip assistant” are normally dismissed as a nuisance after a short period of use. Oracle has therefore designed the Remote Data Capture system to be similarly “process neutral”; the system allows the user to easily find or enter patient data, and then puts all applicable operations on one screen, allowing the user to choose what to do next.

User Interface Overview

The primary user interface for Oracle’s Remote Data Capture system is a spreadsheet, which displays pages of data within the study that the user is authorized to access. Across the top of the spreadsheet is a list of pages within the protocol, optionally tabbed by visit or study phase, and down the left side is a list of patients. Within each cell is an icon which indicates whether there is any data for that page/patient combination; if data does exist, the icon also indicates information about the CRF page, such as whether it contains any discrepancies. Here is a schematic:

	Visit 1		Visit 2		Visit 3		Visit 4	
	1 Demog	2 Vitals	3 PE	4 Hema	5 Chem	6 IncExc	7 ECG	
1 (KLH)								
2 (ABC)								
3								
4								
5								

Patient 2 Page 2 (Vitals)

Date: **1-Nov-99**

Blood Pressure: **80 / 120**

Pulse: **99**

Temperature: **36.9 C**

	Summary	Discrepancies	Verification	Approval	Audit	Help
#	Description		Status			
1	Systolic of 80 is less than Diastolic of 120		New		<input type="button" value="Go"/>	

This example shows the protocol contains seven pages for Visit 1. Patient 1, with initials KLH, has data for pages 1-3, while patient 2 has data for pages 1 and 2. The icon for patient 2 Vitals is red, indicating that it contains discrepancies which require the user’s attention. The “Discrepancies” tab at the bottom shows all discrepancies for the currently-selected CRF page (indicated by the aqua background for the icon). The data for this same page is shown in a window overlaid on the spreadsheet, which the user may move and/or resize as necessary.

The user may click on the other tabs at the bottom of the screen to view other aspects of the currently-selected CRF page. The Summary tab displays information such as the date and user of original creation and last modification, the Verification tab allows the user to perform source data verification (if authorized), the Approval tab allows the user to record an electronic signature (if authorized), and the

Audit tab displays the audit trail of any data changes. The Summary tab is context-sensitive, in that it will display information about the currently-selected object; this includes showing statistics for the current set of data, for the current visit, for the current page number or for the current patient. The patient display includes information such as a list of missing pages for the patient and the projected date for the patient's next visit.

Remote Data Capture Features

Having discussed the benefits of an integrated approach to Remote Data Capture, let's look at some of the specific features of Oracle's offering. This description is for the first release of the system, which is included on the Oracle Clinical 4.0 CDROM, but can also be installed to run against Oracle Clinical 3.2. All that is needed on the user's desktop is a web browser such as Netscape or Microsoft Internet Explorer.

News : On signing on to the system, the user is presented with a news tab, where the content has been provided by the system administrator. The content consists of one or two news items, which can vary according to the study, study site, user id, user role, time window or priority. For each news item there is a headline, some descriptive text, and an optional "More" button which branches to a URL specified by the system administrator. The news system thus allows general or specific broadcasts to the user community, such as notifying users of scheduled downtime, giving notification of study enrollment information or giving access to detailed protocol instructions.

Finding Data : On entry into the system, and at any time while in the system, the user may specify "Selections" which indicate the set of data they wish to access. The selections may include any combination of the following :

- Site
- Patient range
- Visit or page range
- Page status (e.g., "Entry complete", "Batch loaded")
- Discrepancy status
- Verification Status
- Approval Status
- Time window of creation or modification, for data and/or discrepancies
- A specific CRF
- One or two data values within a CRF (e.g., Systolic > 160 or Diastolic > 100)

The current selections are displayed as an English sentence beneath the spreadsheet, to remind the user of the search criteria.

Entering Data : The user may enter new data by clicking on an empty spreadsheet cell and filling in values for the fields as specified in the CRF definition. Typically the user will press the tab key between fields and tab out of the last field to save the data. If a field has a list of possible values behind it, the field will include an icon which the user can press to get the list. The user may enter data which violates the constraints behind the field definition, such as entering a value above the upper bound expected; this

will cause a popup window to be displayed, asking the user to confirm or re-enter the field value. If the value is confirmed, a discrepancy will be raised when the CRF is saved.

Entering Unplanned Data : Occasionally the user may wish to enter data which is not planned according to the study protocol. This may include additional lab tests at a planned visit, or a whole unplanned visit which occurs between two planned visits. The user can cope with these circumstances by requesting an unplanned page within the current visit, or by requesting an unplanned visit. The system automatically generates the appropriate pages within the study schedule, and the user enters data for one or more of these pages. At any time, the user may then choose to show or hide the unplanned pages within the spreadsheet display.

Updating Data : The user may update data by clicking on an existing page icon and changing a field value. This will cause a popup window to be displayed, asking for a reason for the change, with the default reason varying according to the user's role. An optional comment may also be added, and the full audit trail of the change is stored in the database. When the data management staff considers that no further changes should be made to a set of data or an entire patient, then the data may be locked or frozen using standard Oracle Clinical facilities. At this point, site staff would only have read access to the data unless they were given elevated privileges.

Validating Data : Single field failures create discrepancies as the data is saved. The timing for execution of more complex edit checks is controlled by the study designer. The options are to have the check execute as the data is saved, or when the user presses a "Validate" button, or when a data management user runs a process called batch validation. If a CRF contains one or more discrepancies, the corresponding icon will turn red.

Processing Discrepancies : A user may search for all pages containing discrepancies and then review the individual discrepancies by clicking on each page in turn. If the user is authorized to correct the data values causing a discrepancy, that user can click a "Go" button against the discrepancy to get to the appropriate field in the appropriate data entry screen. After correcting one or more data values, providing a reason for change and saving, the discrepancy is re-evaluated either immediately or on the next validation execution, as described above. If the discrepancy no longer occurs it is closed by the system, with an audit trail record created. If the CRF no longer contains any open discrepancies, the icon will turn white.

If the user needs to route the discrepancy to another group, an action may be selected from the drop down list and an appropriate comment may also be added. The drop down list is configured by the system administrator according to the user's role; examples may be "Send to Data Mgt" or "Close - Resolved". The latter option would require a reason for manually closing the discrepancy, and an optional resolution comment. If the discrepancy is routed to another group rather than being closed, then the CRF icon changes from red (requiring the user's attention) to yellow (requiring another group's attention).

Performing Data Verification : A study monitor may indicate that a CRF page has been verified against source data by pressing a button in the "Verification" tab and entering an optional comment. The icon changes to indicate the CRF page has been verified. If the data on the page is subsequently

changed, the icon changes again to indicate that re-verification is necessary. The user may then inspect the audit trail to see what has changed since the initial verification, then re-verify the CRF page. The user may also undo the verification operation if the initial verification was a mistake or is now in doubt. All such verification and undo records are stored in an audit trail and may be browsed online in the Verification tab.

While performing Source Document Verification, the user may choose to automatically move to the next page that requires verification, progressing across the spreadsheet (for the same patient) or progressing down the spreadsheet (for the same page number). In this way, verifying a set of data is quite highly automated.

Performing Electronic Approvals : Electronic approval works exactly the same as source data verification, except that a different privilege level is required. On the first approval, the user is prompted to re-enter the logon password in order to confirm that an unauthorized user hasn't walked up to a logged-on session. Approvals may then be done a page at a time, with auto-progression also used to move to the next page; if the session is inactive for a period (default 10 minutes), then the password must again be re-entered before further approvals can be done.

Reviewing the Audit Trail : At any time the user may review the audit trail for the currently-selected CRF. The user may view the audit trail since CRF creation, since any of the verification or approval timestamps, or since a user-specified date. The audit display shows a row for each field on the CRF that has changed, with the initial value, the current value, and the user, date, time and reason for the change. If there have been intervening changes between the initial value and the current value, the user may drill down to see the complete life history for that data point.

Running Reports : The user may request a formatted report of the patient data, either for all data they are authorized to see in the current study, or for a subset such as a particular patient, or for the selections currently active in the spreadsheet. For each page of data, the report displays a standard CRF header containing information such as the patient and page number as well as tracking information such as the user and date of entry, then it displays the actual data formatted identically to the data entry screens; if the data is wider than the field display width then it will be wrapped on the page.

Data that contributes to an open discrepancy is displayed in bold. The user may also elect to display data which has changed since a certain date in italics; this date may be a fixed date (e.g., flag everything which has changed since 27 Feb 2000), or it may be a variable date (e.g., flag everything which has changed since electronic approval of this page). The report output may be routed to the file system, say as a PDF file, or it may be routed to a printer. Remote sites may use this mechanism to keep a local hardcopy of the patient data on site.

Getting Help : All help and documentation for the system is online. The user may select help from any field in the system and get context-sensitive assistance, or may select a subject from the "Help" tab. The Help tab also includes a description of the different page icons in the system.

Security : Users identify themselves to the system by signing on with a database userid and password. The database userid determines the user's security role, which is selected from a pre-configured set of

values such as “Data manager”, “CRA”, “Investigator”, etc. The user’s name and role are permanently displayed on the window title, along with the current study, site (if selected), date of signon and database name.

The system administrator can specify the studies that each user can access and, optionally, which sites within the study the user can access. At each level, the administrator can specify the set of privileges the user will acquire. The privileges can be any combination of the following:

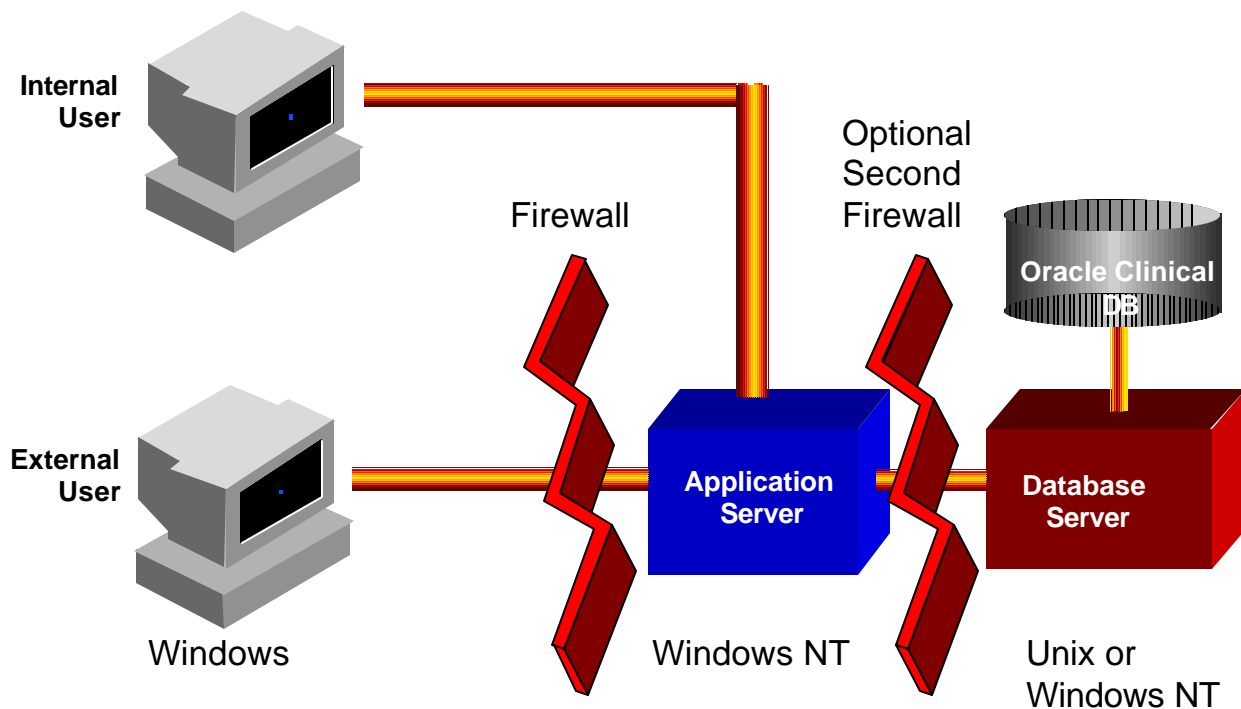
BROWSE	Browse data
UPDATE	Insert/Update/Delete data
UPD_DISCREP	Update discrepancies
BRW_BATCH	Browse batch loaded data
UPD_BATCH	Update batch loaded data
VERIFY	Perform Source Data Verification
APPROVE	Perform electronic approvals

Typical combinations would be as follows :

Data Management	UPDATE, UPD_BATCH
CRA	BROWSE, UPD_DISCREP, VERIFY
Site entry person	UPDATE, BRW_BATCH
Investigator	BROWSE, APPROVE

The system administrator may also control the allowable actions for a discrepancy according to the user’s role. For instance, it may be configured that a site entry person may not manually close a discrepancy, whereas someone in data management is allowed to do this. This allows each company to enforce specific workflows for discrepancy management.

Technical Architecture



The above schematic shows the connection between the user's PC, the application server running the Remote Data Capture system, and the database server running the Oracle8i database. The only software required on the user's PC is Netscape or Microsoft Internet Explorer, plus an Oracle plug-in called Jinitiator, which can be downloaded and installed the first time the user tries to connect to the application server.

If the user is within the company's firewall, the user simply navigates to the URL containing the launch page for Remote Data Capture, and signs on with a database userid and password. If the user is outside the firewall, then one of three methods may be used to establish the connection :

1. By dial-up. This requires the user to dial in to a modem bank located inside the firewall, giving a userid and password in order to establish the dial-in connection. Experience indicates that the application will give adequate performance using a 56k modem.
2. Using a Virtual Private Network (VPN). This establishes a network where only a pre-defined set of nodes may access the application server, although the physical connection is the same as is used for the public Internet. The data may optionally be encrypted over the VPN.
3. Using the Internet. This requires the user to have a connection to the Internet, usually via an internal LAN. The user navigates to a URL which includes a port on the application server which is listening for HTTP connections. Once the connection is established, the user can sign on to Remote Data Capture as usual. The HTTP connection may be made more secure by using HTTPS (a secure sockets connection), which is used in conjunction with a certificate from a third-party vendor such as Verisign.

Compliance with 21CFR Part 11

The Code of Federal Regulations (CFR) Title 21, part 11 covers the management of electronic records and electronic signatures in computer systems. The regulations are designed to ensure that computer systems used for electronic data collection have been developed according to sound practices, that user access to the system is controlled and monitored, that data integrity is preserved on all transactions and that audit trails are maintained. Any Remote Data Capture system needs to comply with 21CFR Part 11.

Oracle considers its Remote Data Capture application to be a closed system, in that “system access is controlled by persons who are responsible for the content of electronic records that are on the system”. Oracle has therefore evaluated the Remote Data Capture application, the back-end functions handled by Oracle Clinical, and the facilities for electronic signatures against the requirements for a closed system. There are a few gray areas in the interpretation of the regulations, such as deciding whether a PDF file of the patient data report constitutes an adequate certified copy of the data at the investigative site, but apart from these areas Oracle believes that the Remote Data Capture system is fully compliant with 21CFR Part 11.

Hardware and Software Specifications

User’s PC :

Pentium class processor, 300MHz or higher
48 megabytes of memory
56k dialup modem or permanent connection to the Internet
Screen resolution of 1024 by 768 or higher
Windows 95, 98 or NT
Netscape 4.5 or higher or Internet Explorer 5.0
Oracle Jinitiator
Adobe Acrobat file viewer for viewing report PDF files

Application Server (specifications for up to 50 simultaneous users) :

Dual 400MHz processor
1 gigabyte of memory
8 gigabytes disk space

Windows NT, service pack 4
Oracle WebDB listener
Oracle Forms 6i Server
Oracle Reports 6i Server
Oracle SQL*Net

Database Server :

Sun Solaris 2.6, HP-UX 11, Compaq Tru-64 4.0G or Windows NT, service pack 4
Standard sizing guidelines (available in a separate document)

Oracle8i RDBMS
Oracle Clinical 3.2 or 4.0
Oracle SQL*Net
Oracle SQL*Plus

Summary

The potential benefits of Remote Data Capture have long been identified. All data is available immediately. The number of data queries that need to be resolved in house are dramatically reduced. Cycle times for those queries that remain can be measured in minutes or hours rather than days or weeks. These in turn can lead to earlier decision-making, improved data quality, and faster database lock.

Lessons have been learned from previous RDC implementations. The deployment and management of hardware and software at investigator sites can be a logistical nightmare, and the problems are magnified as you scale from a pilot implementation to full production use.

While there is clearly a case for moving to a greater number of trials conducted using remote data capture techniques it is important to remember that paper based trials will still be around for many years to come. It is thus vitally important that the data from Remote Data Capture is easily integrated with the traditionally collected data.

Oracle Clinical RDC provides the benefits traditionally associated with Remote Data Capture without any of the drawbacks. There is **no** software to install at the investigator site other than a plug-in to a web browser. There is **no** data stored at the investigator site. It is all stored in a central professionally managed server. There are **no** integration issues. The data is stored in the same database as the data collected by traditional methods.

Further Information

Visit Oracle's pharmaceutical web site at www.oracle.com/industries/pharmaceuticals.