CLUSTER RANDOMIZED CONTROLLED TRIAL FOR IMPROVING PERINATAL CARE IN LATIN AMERICA

STUDY PROTOCOL

Data Management System Specifications

Latin American Center for Perinatology (CLAP)
Pan American Health Organization (PAHO)
World Health Organization (WHO)
Montevideo, Uruguay

Department of Maternal and Child Health
School of Public Health
The University of North Carolina at Chapel Hill
United States

Global Network for Women's and Children's Health Research
National Institute of Child Health and Human Development
National Institutes of Health
United States

Protocol version 8.0
-November, 2002-

1 Introduction

1.1 Purpose

The data management system (DMS) specifications found in this appendix document the DMS for the project "CLUSTER RANDOMIZED CONTROLLED TRIAL FOR IMPROVING PERINATAL CARE IN LATIN AMERICA". The system will be implemented in 26 hospitals (24 study hospitals and 2 pilot hospitals) in Argentina and Uruguay and at CLAP Data Coordinating Center (DCC) in Montevideo, Uruguay. This project is within the Global Network of Women's and Children's Health Research.

2 Data Management System Modules Description

2.1 Overall Description

A description of the data collection and data management activities to be implemented in the study can be found in the study protocol.

In summary, every morning the data manager at the hospital level collects all CLIN1 and ALTA1 forms (Appendix 7) for deliveries and discharges of the previous day and enters the data in the Remote Data Management Subsystem (RDMS) for that hospital. The RDMS is a stand-alone application that runs on a personal computer. Consistency checks are performed during data entry and warnings are displayed. The data is checked to make certain that entered values are acceptable, that all required fields are entered, and that items are consistent with other related items in the database. After completion of the data entry procedure, the system generates Validation Reports (VRs) that must be resolved by the data manager. When resolving VRs, a record of the database update will be kept, identifying information about the person who made the update, the date, changed values, and comments. Reports detailing the status of data collection and validation are available. A data transmission component is available so that data and data VRs from a hospital are transferred to CLAP DCC, where the study database resides.

The Central Data Management Subsystem (CDMS) runs on the central server at CLAP DCC. It consolidates the data collected at all hospitals, manages the collected database, and edits the data. It runs the same reports as the hospitals but on an aggregated database. It will also transmit the data to RTI DCC.

The system provides a secure environment for confidential medical information. Access to the system is limited to authorized individuals. The database is backed up on a regular basis. In addition, information that identifies the patient is stripped from the database when transmitted to CLAP DCC and RTI DCC.

Figure 1 presents a summary of the system data flow.

At the end of the project, the system is release to the public domain and made available to all hospitals in Latin America.

2.2 Security Module

The security module will have two main parts:

- The first part of the security will be at operating system level (Windows 2000). We will create specific users for each hospital, only the people that will work with the system will have access to the computer.
- The second part is implemented at the application level, by defining user profiles. The system will have two user profiles: supervisor and data entry. Users under the data entry profile can complete forms, change a password, run validations, and enter corrections from validation results once validation operation has been completed. The supervisor profile has more privileges than data entry profile. This means that the supervisor profile can do everything that the data entry profile can do; in addition, the

supervisor has privileges for user management, can create users, activate/deactivate user, and reset user's passwords. In case of errors or loss of information, the supervisor can verify the configuration of the system.

At all times the system displays the name of the current user.

2.3 Configuration Module

The system will have a configurations component where the supervisor can modify system variables. For intance, the URL of the central system (IP, Ports,etc), automatic connection periods, expiration of draft documents, etc. All this information will be stored in the database and can be exported/imported as extended mark up language (XML).

2.4 Data Entry Module

The data entry forms have the following features:

- Standard headers identifying the study, the protocol, the site, the form, and the patient.
- Screens that match as closely as possible the data forms developed for the study.
- Allow all entries to be performed with keystrokes and/or with mouse clicks.
- Use check digits to verify correct entry of the patient's identifying number. The user receives a warning if an invalid id is entered.
- Include a range check on all numeric and date fields so as to confirm that responses are valid and logical. The user receives a warning if an invalid response is entered.
- A mechanism that allows the user to override a range error if the value is indeed correct. This provision will require the user to enter a reason for the override.
- Incorporate skip pattern checks at an appropriate point in the data entry process.
- Separate date fields into month, day, and year fields so as to simplify range checks and imputation of partially missing date information.
- Provide four digit fields for years so the centuries may be correctly distinguished.
- Check date fields to ensure they are valid calendar dates.
- At any point during data entry, the user has two options: save the record as draft or save as complete. The data manager can complete draft records in another session. A record can not be in draft state for more than 2 days.
- Records saved as complete can be modified only through VRs (see validation module).
- Provide the user with an option to print a copy of a completed or partially completed form.

2.5 Data Validation Module:

After completion of the data entry procedure, the user runs the validation module. The user can select the set of records to validate. Records saved as Draft can not be validated at this level.

The system generates VRs that must be resolved by the data manager. The RDMS has a module for VRs management where reports can be sorted, listed, printed, and resolved.

When resolving VRs, a record of the update will be kept, identifying information about the person who made the update, the date of the update, the value of the field prior to the change, the value after the change, and any relevant comment made by the data manager.

The data manager can generate new VRs to allow edits not considered by the validation engine.

All corrections are stored in a separate table of the database. The original record is not modified.

When the user logs into the DMS, a pop-up reminder is presented showing the number of pending VRs, according to type of report and days pending resolution. This report will also be accessible from the validation module.

All VRs are sent to CLAP DCC on a daily basis (see transmission component), where the state of the validation process is reviewed.

Further validation routines will be performed by data supervisors at CLAP DCC (detection of outliers, consistency with previous data, etc.). VRs can be generated by CLAP DCC and sent back to the hospital.

2.6 Data Transmission and Backup Module.

The final specification for this module will be defined in collaboration with RTI.

We are reviewing two options:

- BLAST data pump: This product is similar to the FTP protocol, because it transfers files between systems. This product has many advantages: it is robust and can be run in many platforms and over many protocols. However, the communication is non-synchronized, is not standard, and requires licenses.
- XML messaging system: This option is also robust and run in many platforms, but it has a number of additional advantages over BLAST. XML does not require licenses and will allow CLAP to release the software free of charge. Communication between remote and central system should be synchronized. This allows both systems to keep the information about each other on-line, and the system transmits small pieces of data. For instance, the system will establish an Internet connection with the central system and will send only the registers marked as “to send.” Once the data is transmitted, the central system responds “OK” or “Error” in transmission. In this last case, the data will be sent again. The system will send and receive data as XML. XML is a document like HTML documents, but it has important differences. An XML document typically consists of markup and character data. The markup information conveys the meaning behind the document contents and is represented through tags and other XML elements, while character data represents the actual content. In the fallowing example, the <author> tag indicates that “Joe Bloggs” represents an author’s name:

  <author>Joe Bloggs</author>

In the system, the user can always start the transmission process manually, but the system will do it automatically too, establishing periods of connections. If the system cannot connect in “n” times with the central system, it will communicate to the user that there is a problem with the connection. The user (data manager or supervisor) must check the physical connection (wire) or see the configurations of the system. In addition, the system will check the date of last successful transmission, and if more than “n” days have elapsed from the last

connection, the system, once again, will display an alert or warning. In this system, the server has detail information of the state of the remote computer. Finally, there are many ways to transfer XML messages over networks, the system will use HTTPS protocol or socket connections.

Include an option for database backup.

2.7 User Management and Audit Module

The supervisor can create, activate/deactivate, or reset the password for a user. When there is a new user, the supervisor will record the information about the new user into the system (first and second name, age, etc), then the system will set a default password for that user (for instance: userxx or 123456 or just “password)”.

Once the user enters to system, he needs to change his password. If the user no longer works in the system, for any reason, the supervisor can “delete” (deactivate) the user (no one can use that username again; it is very important to keep a true recording of all activities of the remote system). This delete is at the logical level in the database and the supervisor can not do a physical delete of that user.

The supervisor can reset users’ passwords. The users’ passwords are stored with an encrypted algorithm so no one (users and supervisor) will see the real password. Each keyed password is encrypted and compares with the encrypted version of the database; if they match, the access is grant, if not, the user will have two (or more) attempts, and then the system will block the user for security reasons.

All activities that a user performs will be logged by the system. The operations that the system will track are: Logon, Logout, documents validated, process of validation, modifications of system configurations, date of last successful connections, and other things.

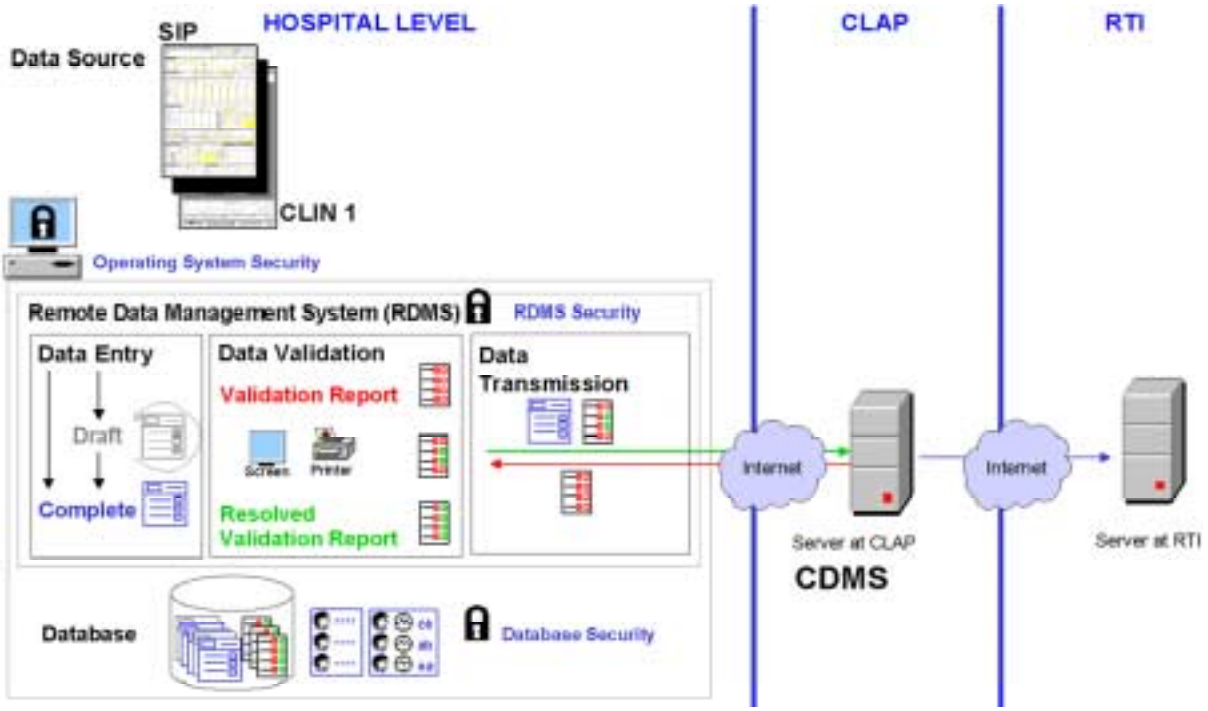
3 Design Issues

The system will be developed using software modules. Each module will be developed as an object, and the object will have an interface to communicate to other objects. Using this strategy allows for the modifications of objects without affecting the general functionality of the system, as long as the interface remains unmodified. This is the basis in Object Oriented Architecture.

The programming language that we will use is C# in the dotNet platform, and we will use this software for all subsites.

The system will adhere to Microsoft Graphical User Interface standards (GUI).

The remotes systems will not be a multiuser environment.

The system will keep the forms, queries, lookup tables, and reports in a separate database from the actual data and will keep the information about patients in a separate database.

4 Figure 1. Data Flow Diagram