PART III: STUDY MANAGEMENT AND SUPPORT

13. INFORMATION MANAGEMENT SYSTEM (IMS)

13.1 Introduction

The Information Management System (IMS) is integral to the National Children's Study. The IMS houses all NCS-related information and serves investigators and the public throughout the Study lifecycle. At the earliest stages of the Study (study design, recruitment, and enrollment of expectant mothers and the Study launch), the IMS records and tracks enrollment, personal information, and informed consent. Through pregnancy, birth, childhood, and adolescence, it supports the tracking of participants, collection of data, report of findings, and incentive management. As Study Centers and data collectors collect biological data and samples, physical measures, environmental data and samples, and questionnaire and assessment data, and as laboratories analyze those specimens, the IMS records, transforms, analyzes, reports on, and protects the information. The IMS also maintains information regarding the location and disposition of physical samples and the results of the sample analysis. To facilitate data gathering, the IMS assists in scheduling visits, including generation of visit reminders to participants and schedules of upcoming data activities for data collectors. Prior to going into the field, the data collectors upload all data needed to conduct interviews and assessments, including participant and schedule information.

Critical features of the IMS include its ability to collect, to store, and to report on the data during the Study as well as to store and report on the data after the Study is complete. To reduce the risks associated with data collection, storage, and reporting, data storage for the Study will be centralized in the IMS at the Coordinating Center. Data are gathered through multiple means (such as laptop-based survey instruments, Web-based interfaces, and measurement devices) and are electronically sent to or entered into the IMS. Backup and protection of all data are guaranteed by the centralized storage. The IMS supports centralized, uniform, high quality data collection and analysis activities for the Study. The IMS also supports uniform and consistent participant de-identification and strong controls over re-identification, as well as producing investigator-specific data sets.

Since critical Study activities are supported centrally, the IMS maintains continuous "24x7" Study operations. It incorporates state-of-the-art redundancy, fault tolerance, and disaster recovery mechanisms to ensure that operations can continue if hardware, software, or communications fail. The majority of IMS functionality is accessible to the Study Centers through an Internet browser over a secure network. Other functionality is accessible through disconnected data collection devices (e.g., laptop computers used by field data collectors).

Data collection in the home or at other field locations utilizes laptop computers and, occasionally, environmental sampling devices. This collected information is synchronized with the central database when the data collector is able to log into the Coordinating Center (either remotely or through a direct connection in a Study facility).

Clinical event data collection regarding such data as diagnoses, interventions, etc., which occur over time, constitute important outcomes and exposures to incorporate into the participant data base. Methodologies are being studied to facilitate obtaining these data during the Study from disparate sources such as primary care physicians, specialty consultants, hospitals, emergency rooms, and public health clinics.

13.2 Security and Privacy

Security and privacy are factored into every aspect of the IMS design. Security includes protection of sensitive data from corruption, theft, tampering, or unauthorized use, as well as protection from loss or corruption due to internal problems (e.g., a hardware or software failure) or external forces (e.g., a natural disaster). Privacy restricts access of sensitive information to only those individuals who are authorized to use or to view such data. De-identification of data—separating potentially identifying personal information from the actual participant data—is one aspect of privacy.

The IMS is hosted on a number of dedicated servers in a secure facility where physical access to the servers is restricted only to authorized personnel. The Study data are stored and managed in a secure network environment that is protected by continuously updated firewall, anti-virus, and anti-intrusion hardware and software. Systems are actively monitored to detect and block any attempt at intrusion or "hacking." Secure network connections are established between the Coordinating Center and external entities (e.g. Study Centers, labs, and repositories) to ensure data are not compromised during transmission. All data are encrypted during transmission and upon storage.

The IMS complies with various policies and regulations, such as the Health Insurance Portability and Accountability Act (HIPAA), to protect the privacy of the participants. Even Study staff, such as the Coordinating Center data managers and analysts, are not able to associate Study data with actual participant identities except under strictly defined conditions. To fulfill this mandate, the IMS employs a second layer of security specially designed to segregate participants' personal identifying data (PID) from the rest of the data. PID is stored in specially encrypted databases on servers that are physically separate from the main database servers. User IDs, passwords, and "digital certificates" allow access to PID only by authorized individuals and from authorized access points. The password/digital certificate system may also be augmented by a biometric identification technology such as thumbprint scanning to guarantee that any request for PID is genuine and coming from an authorized user.

13.3 Architecture/Framework

Since the Study will last more than 20 years, the IMS is designed with the ability to grow with the Study and to adapt with the evolution of technology. The IMS framework allows reusing existing applications and systems while accommodating future technology expansion. The framework accomplishes this by focusing on interoperability and component-based architecture.

Interoperability is defined as the ability of different types of computers, networks, operating systems, and applications to work together effectively to exchange information in a useful and meaningful manner. Interoperability requires not only the ability to transfer data, but a common understanding of what those data mean. The IMS supports interoperability with other systems (e.g., external databases with relevant data) by including multiple methods of transferring data between systems and by the use of industry standards to define not only the syntax but also the meaning of the data. Leveraging standards for integration enables the IMS to be flexible when future technology changes are implemented.

The IMS is a component-based architecture in which "components" (e.g. system building blocks) are responsible for specified functions. These components have well-defined interfaces. This approach supports later replacement of a component with newer or alternate versions that enhance functionality or incorporate new technology (in a "plug and play" manner). The result is a scalable IMS adaptable to the Study's long-term goals.