(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
12 April 2001 (12.04.2001)

PCT

(10) International Publication Number
WO 01/25938 A1

(51) International Patent Classification⁷: G06F 13/00, 13/38, 15/16, 15/17, H04Q 7/00, 7/20

(21) International Application Number: PCT/US00/27020

(22) International Filing Date: 2 October 2000 (02.10.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/157,357 1 October 1999 (01.10.1999) US

(71) Applicant: CB TECHNOLOGIES, INC. [US/US]; 1487 Dunwoody Drive, Glenloch Corporate Campus, West Chester, PA 19380-1478 (US).

(72) Inventor: HUME, Samuel, W.; 1179 Meredith Lane, Chester Springs, PA 19425 (US).

(74) Agent: DICHTER, Eric, A.; Wolf, Block, Schorr and Solis-Cohen LLP, 22nd Floor, 1650 Arch Street, Philadelphia, PA 19103-2097 (US).

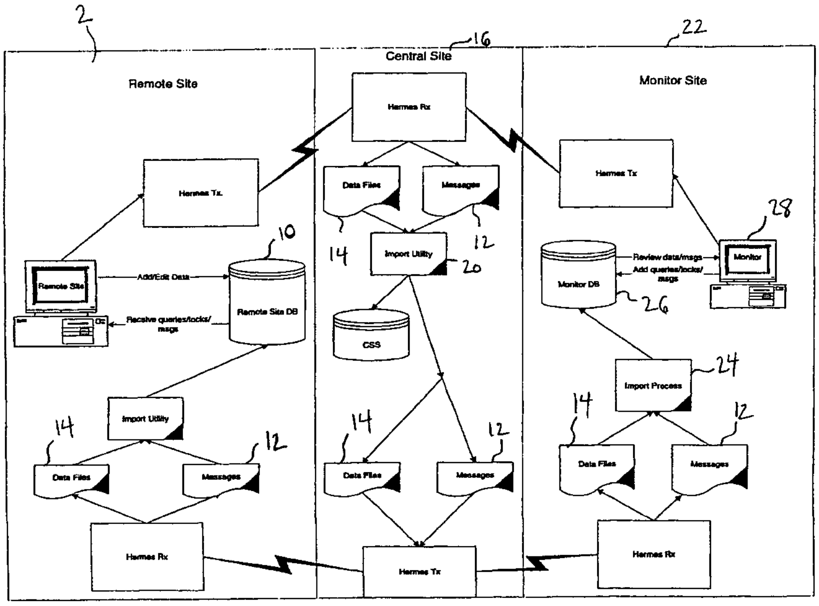
(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *With international search report.*

[Continued on next page]

(54) Title: METHOD AND SYSTEM FOR ELECTRONIC DATA CAPTURE



(57) Abstract: A method, system, and computer-readable medium for collecting and processing electronic data connects one or more remote user sites (2) where data is entered, processed, and transmitted with a central site system (16) for receiving and processing the transmitted data by a computer network connection. In addition, the central site system (16) is connected to one or more remote monitors (22) for evaluating and possibly altering the data. The method, system, and computer-readable medium allow for entry and processing of data if a network connection cannot be established and enable continuous, real-time entry of data at the remote user site (2) without significantly affecting performance.

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

METHOD AND SYSTEM FOR ELECTRONIC DATA CAPTURE

Field of the Invention

The present invention is directed to electronic data collection. More particularly, the present invention is directed to a hybrid system and method of electronic data capture in which data is entered from a remote site and it is processed and transmitted to a central site for further utilization of the data.

Background of the Invention

In many different applications, e.g., the pharmaceutical industry, the collection, cleaning, and management of data obtained from multiple sources are essential and act as a bottleneck in accomplishing desired goals. Even in this electronic age, paper collection of data is still used for certain applications, e.g., using written case reports for collection of data for pharmaceutical clinical trials. Such paper collection is slow; it is difficult to reproduce, distribute, and evaluate paper collected data; and it is onerous to store such data.

Thus, in some instances, an individual will enter data electronically using a computer having software which processes the data. This data collection method is limited in that data is entered, processed, and used at only the individual computer. Thus, electronic data collection and processing in connection with a computer network, such as the Internet, would progress beyond the individual computer to enable multiple users to enter information and use such information. Such a configuration has been used in many instances, especially with respect to Internet transactions.

In a typical situation, an Internet address/world wide web page provides a collection site for data to be entered into a central database. The database is located behind a web server and can be accessed by a multitude of users in order to submit information (data). Nevertheless, these Internet-based data collection systems do not allow users to enter data off-line such that Internet performance, e.g., connection speed, connection reliability, etc., becomes a variable in the ability to enter data and a slow connection speed can significantly hinder the entry and processing of the data.

Also, these Internet-based systems transmit to the central database data forms facilitating entry of data to be collected along with the data itself, which creates larger pieces of data to transfer and makes transmission further subject to Internet or other network connection limitations. Such limitations may include a slow response time in entering the data while data is being transmitted.

Such Internet-based systems use secure sockets layer (SSL) as a security measure, not the other security devices used in Internet communication. Further, communications with such systems may not be real-time, the response time is directly correlated with connection speed such that low-bandwidth and/or wireless networks may not be used, they are not able to work off-line, and they may not be fault tolerant. In addition, the completeness and acceptability of the data is only judged once the data filters through to the web server such that unnecessary, unuseful data is transmitted.

Further, there are security concerns with data collected through the Internet. For example, hackers have brought down sites for extended periods of time and have stolen and/or corrupted data from various sites. In addition, computer viruses can be spread through use of connections, e.g., use of e-mail communications, to the Internet. Those systems that work both off and on-line, such as Microsoft Money® (a product of Microsoft Corp.) and Quicken® (a product of Intuit Corp.), are not designed to capture data from a remote user and they have certain delineated functions designated for on-line and others designated for off-line.

In the area of clinical trials of pharmaceutical products, data collection from many sources and analysis of such data is necessary for assessing the efficacy of such products. Thus, there exists a need for a secure, efficient method of electronic data capture and processing which overcomes the shortcomings of conventional methods.

Summary of the Invention

In its most general form, the present invention comprises a method, system, and computer-readable medium proceeding in real-time for collecting and processing electronic data. In one embodiment, the method, system, and computer-readable medium comprise entering data at a remote user site, comparing the data to preselected characteristics for each specific type of data to determine acceptability of the data, transmitting an acknowledgement of the acceptability of the data, storing the data at the remote user site, converting the data into packets, transmitting the data in the form of data packets to a central site system, and storing the transmitted data at the central site system.

In one embodiment, the method comprises using the transmission control protocol/internet protocol, hypertext transfer protocol tunneling using port 80, to transmit the data packets from the remote user site to the central site system with a computer network connection. In a further embodiment of the method, each data packet is encrypted before each data packet is transmitted and the central site system decrypts the data packets after the packets are transmitted; the central site system identifies the data for grouping with the corresponding fields. In yet another embodiment of the method of the present invention, the data stored at the remote user site is synchronized with the data stored at the central site system.

In an additional embodiment of the method of the present invention, if the network connection between the remote user site and the central site system is intermittent, the remote user site detects if a network connection is present and continuously attempts to establish a network connection. In yet a further embodiment of the method of the present invention, data at the central site system is transmitted to remote monitors who evaluate the data by applying comments to the data or locking the data.

In one embodiment, the system comprises software adapted to use the transmission control protocol/internet protocol, hypertext transfer protocol tunneling using port 80, to transmit the data packets from the remote user site to the central site system with a computer network connection. In a further embodiment of the system, the software is adapted to encrypt each data packet before each data packet is transmitted, decrypt the data packets after the packets are transmitted, and identify the data for grouping with the corresponding fields. In yet another embodiment of the system of the present invention, the software is adapted to compare data stored at the remote user site with the corresponding data stored at the central site system to synchronize the data.

In an additional embodiment of the system of the present invention, if the network connection between the remote user site and the central site system is intermittent, the software is adapted to detect if a network connection is present and continuously attempt to establish a network connection. In yet a further embodiment of the system of the present invention, the software is adapted to transmit data at the central site system to remote monitors who evaluate the data by applying comments to the data or locking the data.

In a further embodiment, the computer-readable medium comprises using the transmission control protocol/internet protocol, hypertext transfer protocol tunneling using port 80, to transmit the data packets from the remote user site to the central site system with a computer network connection. In a further embodiment of the computer-readable medium, each data packet is encrypted before each data packet is transmitted and the central site system decrypts the data packets after the packets are transmitted; the central site system identifies the data for grouping with the corresponding fields. In yet another embodiment of the computer-readable medium of the present invention, the data stored at the remote user site is synchronized with the data stored at the central site system.

In an additional embodiment of the computer-readable medium of the present invention, if the network connection between the remote user site and the central site system is intermittent, the remote user site detects if a network connection is present and continuously attempts to establish a network connection. In yet a further embodiment of the computer-readable medium of the present invention, data at the central site system is transmitted to remote monitors who evaluate the data by applying comments to the data or locking the data.

In one embodiment, the present invention comprises a method (also applicable to a system and computer-readable medium of the present invention) for synchronizing electronic data at a remote user site with data at a central site system comprising comparing data entered into and stored at a remote user site with data transmitted to and stored at a central site system by transmitting data between the remote user site and the central site system via a network connection. In another embodiment of such method, the data is transmitted using a transmission control protocol/internet protocol. In a further embodiment of such method, if the network connection is intermittent, the remote user site continuously attempts to form a network connection to transmit the data.

It is to be understood that both the foregoing general description and the following detailed description are exemplary, but are not restrictive of the invention.

Brief Description of the Drawings

The present invention is best understood from the following detailed description when read in connection with the accompanying drawings. It is emphasized that, according to common practice, the various features of the drawings are not to scale, rather, the dimensions of the various features are arbitrarily expanded or reduced for clarity. Included in the drawings are the following figures:

FIG. 1 is a schematic illustration of the data collection and processing method and system according to an embodiment of the present invention; and

FIG. 2 is a schematic illustration of the data collection and processing method and system according to an embodiment of the present invention including a remote user, central site system, and a monitor.

Detailed Description of the Invention

The present invention comprises a method, system, and computer-readable medium for collecting and processing electronic data entered into a computer at a remote user site. For the purposes of the present invention, remote user site is a client site or remote node exchanging data with a central server used by one or more remote users. The method, system, and computer-readable medium enable a remote user to be in communication with a central site system and to enter data with a computer at a remote user site. For the purposes of the present invention, a central site system is a central computer server component which manages the various remote site connections and authenticates the connecting sites. The remote user site, by definition, does not necessarily have a continuous connection with the central site system. The data is evaluated according to preselected (known) characteristics of the type of data being entered at the remote user site and, if the content of the data is acceptable for the type of data being entered, it then may be stored at the remote user computer. If the data is not acceptable, the remote user is notified and is allowed to reenter data corresponding to the data field(s) deemed unacceptable.

If a network connection is established between the remote user site and the central site system, the data can then be transmitted to the central site system in the form of data packets where the data can be further processed. In this case, the transmission of data proceeds in real-time while the user enters further data. For the purposes of the present invention, real-time is no more than one second response time for data entry and notification functions of the system. In fact, because the data is packaged and then transmitted, the transmission should not impact the computer performance in processing the entered data significantly, if at all.

If a network connection cannot be established or is intermittent, the system allows for the entry of data at the remote user site; and the data is evaluated for acceptability and stored at the remote user site with no data being transmitted to the central site system. Meanwhile, the system continuously attempts to establish or restore a network connection and can store the information and corresponding notification of the failure of a data transmission even in the event of a power failure. If the connection is not able to be established, such as after several failed attempts to connect, the system will work off-line and data entered at the remote user site is placed on a diskette which may be sent to the central site system by non-electronic means. Further, the system is fault-tolerant such that errors occurring in the system, e.g., failure to connect with the central site system, do not prevent a remote user from entering or altering data.

The computer-readable medium, i.e., software, used with the system and method of the present invention is a browser-based system having a communications component that functions from the central site system and remote user sites. The computer-readable medium of the present invention may comprise a conventional commercial browser product, such as Internet Explorer (a product of Microsoft Corp.), modified to limit its browsing capabilities to the data collection and processing system of the present invention, i.e., not for general Internet access, and provide the necessary communications architecture, including reception, processing, and transmission of data. In fact, the communications system is decoupled from the browsing system. The computer-readable medium is on the central site system and the remote user site computers.

The system (the description of which is also applicable to the method and computer-readable medium), as shown in FIGS. 1 and 2, receives data entered into a computer at a remote user site 2 using software 6 which is adapted to receive data entered by a user, process the data, and communicate with a central site system via a computer network 4. The remote user must log onto the system with a form of identification, such as a user identification (userid) and a password which can be changed periodically for added security.

The security features of the present invention limit system access to authorized users only. In addition to the userid and password required for access to the system, access to features within the system is limited based on a user's specific rights. Access rights are determined by the role(s) to which a user is assigned.

Administrators of the system can create a role, assign rights to the role, and add individual users to each group. In this instance, access rights are correlated with the particular system features to which a group will be given access. For example, remote monitors (as described in more detail below) have the right to generate data queries on fields and site coordinators of remote user sites or remote monitors do not. Nevertheless, site coordinators may have the rights to add and/or edit data at the site while monitors typically do not have this right.

The access control features include:

(1) gaining access to the system by entering a legitimate userid and password;

(2) creating a 2-factor authentication mechanism by adding operating system and/or virtual private network (VPN) (or other network type) authentication;

(3) users are assigned roles within the system - system administrators create roles by assigning them a specified set of system rights;

(4) administrators can create as many roles as necessary and can assign any variety of rights to the given roles;

(5) administrators can add, remove, and/or modify users and roles centrally (at the central site system) or at the remote user site;

(6) administrators can disable individual user accounts, or remove rights at the role level;

(7) administrators can set a date after which all editing privileges are removed from all users, regardless of role, to enable remote monitors to expeditiously work toward database lockdown while controlling changes at the site;

(8) once a user is granted access to the system, that user only has the ability to perform those functions for which that user's role was granted access;

(9) administrators can set the maximum allowed idle time within the system, i.e., if the system remains idle for a period of time exceeding this parameter, the system automatically executes a logoff;

(10) administrators may also set a maximum allowable number of failed logon attempts, i.e., if the user exceeds this number of attempts, the system is locked from all logons until the administrator resets the system;

(11) using an operating system (OS) that is securable, i.e., a 32-bit Windows operating system, such as Windows 95, Windows 98, Windows NT 4.0, or Windows 2000 Professional (products of Microsoft Corp.), to add an extra layer of security to the application at the site;

(12) all data stored on the local system's hard drive, including the local database, is encrypted using the RC-4, TDEA (triple DES) (SHA-1), or other compatible algorithms;

(13) the system transmits non-contextual data, i.e., the data packets contain data without references to their inherent meaning;

(14) unlike traditional, thin-client Web systems, the data is not transmitted within the HTML CRF form;

(15) the meta data maintained within the system contains the keys to interpreting the data packets such that, because there are not direct links to patient information, a patient's privacy in maintained. In the worst-case scenario, data from only one site would be exposed. If all attempts to limit access to the system are violated, the cracked site only exposes its own data. Specifically, remote user sites do not have access to the entire central site system database (the central site system database is not on-line) and, therefore, one remote user site cannot compromise the entire study database.

Once a remote user is logged on and enters data, the characteristics of the data entered are compared to known, i.e., preselected, characteristics (or rules) 8 for a specific type of data to be collected, e.g., the results of pharmaceutical clinical trials, to determine if the data is acceptable for the specific type of data. The user at the remote user site 2 is then notified of the acceptability or lack of acceptability from a message 12 or other display transmitted at the remote user site 2. For example, if a user reporting the results of a clinical trial on a human subject reports the body temperature of that subject to be 80°F that data will be deemed unacceptable as non-credible for a living individual. In this example, the preselected characteristic is human body temperature and 80°F falls outside of the acceptable range. The user will be notified of the non-acceptance of the value and will be allowed to change that value. The new value entered will then be judged for acceptability.

After the entered data is deemed acceptable, it is stored at the remote user site using a storage device 10, such as a computer hard drive. The data is then converted into data packets 14, comprising data elements and separators, of various sizes to facilitate transmission of the data to the central site system 16 over a computer network 4, such as by a connection to the Internet 4. Unlike conventional Internet data collection methods, in the present invention, the data is mostly transmitted from the remote user site 2 to the central site system 16 without transfer of the forms which facilitate entry of such data. An exception to this is that forms are transmitted between the remote user site and the central site system when they are updated. In addition, no meta data is included within the data packets such that data can only be recognized by users of the system; the system that receives the data can then recognize and group the data with the proper fields. This allows for the formation and transmission of smaller packets of data thus enabling compatibility of the present invention with multiple types of network connections of varying bandwidths and connection speeds.

Preferably, the size of the data packets is 64,000 bytes or less to allow efficient transfer of the data over a computer network. Such a network may be the Internet, a local area network, a wide area network, an intranet, a dial up connection, a virtual private network, or any combination of these network architectures. In addition, the connection may be provided through a dedicated network line, such as a T1 or ISDN line, or a dial-up connection, and the network may be a wireless network. Where applicable, dial-up connections are established via a remote access server through a local Internet service provider (ISP). Thus, the scripts available from the ISP may be used to establish dial-up connectivity worldwide and the system is compatible with different international telecommunications environments.

The data packets are encrypted with a digital signature, such as a 128-bit secured hash algorithm or other compatible encryption technique. The algorithm is set up such that the hash total is generated using information in the data packet and additional private information known only to the site and central site system such that it cannot be decrypted with information only in the data packet. Other types of data encryption algorithms that may be used include, without limitation, RC-4/40-bit, DES/56 bit, and Triple DES/168-bit (112-bit effective). The encryption scheme used with each of these algorithms is private key encryption, such as PGP (Pretty Good Privacy), with each remote user site and the central site system being the only ones with enough information to decrypt the data. The nature of the data packets, i.e., no meta data contained within them, also enhances security in that they contain information that is out of context to non-users of the present system, although identifiable by the present system. This encryption makes the data packets more secure than with use of SSL because, unlike with SSL, the data is encrypted before leaving the application (program). In addition to this encryption, the system encrypts all data stored on the local hard drive of the remote user site.

The encrypted data packets are transmitted from the remote user site 2 to the central site system 16 via a transmission control protocol/internet protocol 18, such as hypertext transfer protocol tunneling using port 80 without opening another port. Alternative ports may be used as long as they are compatible with the present system. This type of protocol allows the data packets to be transmitted ensuring the integrity and authenticity of the data. In addition, conventional Internet security measures, such as firewalls, routers, packet filters, and proxy servers, and very little or no configuration change for these measures would be necessary. The system may be configured to have multiple communications variations, i.e., from a user's regular office or from an additional site if the user is traveling. Also, third party transmission mechanisms, such as pcAnywhere or Xcellenet, can be used with the system. The system capability is greater than it would appear with the hardware coupled to the system, e.g., it can handle over 1,000 concurrent users with hardware generally used for small networks.

Once transmission is attempted, the user at the remote user site is then notified of the status of the data transmission. This 2-way communication is accomplished using TCP sockets and the WinSock application. The notification does not significantly interfere with the performance of the data entry being performed. In fact, the system has a consistent response time of one second or less.

At the central site system 16, the computer-readable medium attempts to authenticate packets transmitted from the remote user site 2. Data packets are filtered out if they are not authenticated for one of the following reasons:

(1) the communications system fails to properly decrypt the packet (this decryption will fail if the packet has been changed in any way);

(2) the digital signature in the packet cannot be properly re-generated. The SHA-1 hash algorithm (with encryption) is used to generate the digital signature. If

any bit in the packet is altered, the digital signature will not match. Information used to generate the digital signature is added by the receiving system (this information is not included in the transmitted packet). This information is generated by the system based on the remote user site that transmitted the data. If the site was spoofed, the data added to the packet and the algorithm used to manipulate this data will result in a digital signature that will not match.

(3) the key information in the header does not correspond to a remote user site that is registered for the applicable study; or

(4) the format of the data packet has been altered. A redundant check is also available to ensure that the packet has not been altered since it was created.

The remote user site employs the same mechanisms for filtering packets sent from the central site system to the remote user site.

The system of the present invention also features many capabilities to limit access to information transmitted over the network. Network security features include:

(1) the system filters packets that do not pass authentication;

(2) the system employs validated data and parameter checks to ensure that it is not susceptible to buffer overflow attacks;

(3) the system does not use system shells or other launch executables and, thus, it is not possible to manipulate it into launching a program that would give an attacker access to the system, as seen in many common gateway interface (CGI) related attacks;

(4) administrators can limit traffic to and from the remote user, i.e., client, system to specified IP addresses;

(5) the system does not maintain, or cache, connections to a remote user site once the data packets are transmitted and, thus, it is not possible for attackers to maintain a "hacker's bridge" into another system;

(6) because it does not support other protocols, such as SMTP, the system is not susceptible to e-mail viruses and worms;

(7) because the system transmits data packets, and does not transmit executables, it is not susceptible to traditional virus attacks (in the sense that it will not download or communicate viruses);

(8) the system's multi-tiered authentication mechanism makes it very difficult to spoof. Those packets that are spoofed fail at least one level of authentication and are filtered. From a network security standpoint, IP spoofing is difficult to defend against, however, spoofed data will get filtered, prior to making it into the database;

(9) DOS and especially DDOS attacks are difficult to defend against. There are network security measures that can be initiated to limit the effectiveness of such attacks. Fortunately, the remote users are minimally effected by DOS/DDOS attacks, since they operate with consistent performance regardless of the server's ability to process the information. Furthermore, even if the server is taken off-line, the sites can continue their work uninterrupted; and

(10) the system does not use cookies.

Once the central site system is able to decrypt the data packets using technology compatible with the encryption technique used. The data packets are identified 20 based on their content and the fields which they contain. The software at the central site system recognizes such fields and can place the data in the proper organization based on the field information. Once the data packets are properly organized into data related to certain fields and forms, the data is stored at the central site system in a server hard drive 22 or other compatible data storage device. It can also be encrypted before being stored at the central site system.

After the data is transmitted to the central site system, decrypted, organized, and stored at the central site system, the data stored at the remote user site is compared to the data stored at the central site system to synchronize the data. The synchronization can be carried out continuously during a data entry session or at distinct times, such as at the start of or end of a session. If the data is detected to be unsynchronized at the remote user site and the central system site, data can be prepared for retransmission and retransmitted from the remote user site to the central site system and then rechecked for synchronization.

Another feature of the present invention, as depicted in FIG. 2, is that the data at the central site system can be transmitted to other users having a network connection to the central site system. These users, i.e., remote monitors 28, can review the data for compliance with proper standards according to the type of data that is being entered for a specific purpose.

In the case of data for pharmaceutical clinical trials, there are often multiple reviewer groups that evaluate the clinical data and provide comment on such data. The reviewer evaluation and comment often lead to recharacterization of the data, reentry of the data, and/or repetition of experimental procedures yielding the data. Alternatively, if the data is deemed to be acceptable for submission to the Food and Drug Administration or any other regulatory agency or group, the data can be locked, i.e., maintained in its present form.

The central site system converts the data to be sent to a remote monitor site 22 into data packets, in a process substantially similar to that used in preparing for transmission of the data from the remote user site to the central site system. The data packets are then encrypted according to the methods described above and transmitted to the remote monitor site. The data packets are received and decrypted by the remote monitor sites with remote monitors being notified of the status of transmission. The data from the data packets is then organized into data related to certain fields and forms based on relevant fields in the data packets 24. Such fields and forms are recognized at the remote monitor site. The organized data is stored at the remote monitor site in a computer hard drive 26 or other compatible data storage device.

Once the data is organized and stored at the remote monitor site 22, the data is substantively evaluated by the remote monitors. The remote monitors can provide feedback by applying comments to the data or, if the data is suitable for submission, further follow up experimentation, or any other purpose related to the use of the data, the data can be locked by the remote monitors such that no further changes to such data can be made.

Once the data is altered, i.e., by comment or locking, it is converted into data packets and encrypted, both of which are as described above with respect to transmission of data between the remote user site and the central site system. The encrypted data packets are then transmitted to the central site system, the status of which is provided to the remote user site. At the central site system, the altered data is decrypted, organized, and stored at the central site system where such data can be accessed by other remote monitors or the remote user(s) who initially entered the data. Further, the central site system keeps a log of all transmission activity going to and from the central site system.

The accountability of the system of the present invention insures that all system actions are tracked and available for review. Due to the regulatory requirements placed on an electronic data capture (EDC) system for clinical trials and possible requirement for other applications involving data capture, the system includes a complete audit trail of all data captured by the system, and all actions performed on the system. Users with appropriate system rights may view and even print the audit trail. Nevertheless, it is not possible to alter or edit the audit trail. All data edits and changes are captured by the existing audit trail. Additional audit trail features include:

(1) the system maintains a full audit trail for all data entered into the system capturing who, when, what, and why changes were made;

(2) the system maintains a full record of every data transmission packet created, the records included in the packet, the transmission status, when the packet was created, when the packet was transmitted, and how many times the packet was retransmitted. Each packet is assigned a sequential ID. Gaps in the sequence indicate missing packets. Thus, the system should account for every packet;

(3) every data packet received is logged and stored in a file as part of the overall audit trail;

(4) the system tracks every action selected by the users. For example, every time a user edits a form or saves the edits, the request for the action is saved in a log;

(5) a complete audit of all mid-study updates is maintained at the central site system and at the applicable remote user site;

(6) the system tracks every system logon and logoff;

(7) every invalid logon attempt is tracked;

(8) each time the system is not properly shutdown, an audit trail is created and reported to the administrator;

(9) each packet that is filtered out, at the central site system or at the applicable remote user site, due to failed authentication is logged and saved in a suspense directory;

(10) the system includes very rudimentary intrusion detection capabilities. These features can be complemented by a network level intrusion detection system, including system access logs;

(11) the system uses reconciliation mechanisms to actively assert that data synchronization is correct and complete; and

(12) the system employs redundant synchronization engines to ensure that the clinical data at the site is properly replicated at the central site system.

The accuracy of the system allows for verification of the integrity of the data. The system's hybrid architecture uses the following mechanisms to verify the accuracy and integrity of the data that is transmitted from a remote user site to the central site system and back out to other remote user sites:

(1) the integrity of the data transmitted to the central site system can be verified using encryption/decryption with the private key, a digital signature, regenerating the hash total, and a cyclical redundancy check (CRC). If any of the checks listed below fail, then the packet is considered corrupt and is filtered out;

(2) because the packet, including the digital signature and the CRC, is permanently stored as part of the audit trail, the integrity of the data can be re-verified

(using a utility) at any point in time to satisfy an audit or other accuracy or authenticity challenge;

(3) using a VPN, containing a public key infrastructure (PKI) encryption/authentication system, in addition to the system's organic security mechanism adds an additional layer of verification;

(4) the system uses other data structure tests that ensure the packet is well-formed. If the packet has not been formatted according to the internal rules for transmitting the data, then the source of the packet is considered dubious;

(5) the system uses data redundancy to provide a reconciliation mechanism. Reconciliation is a key component of ongoing system auditing and maintenance. The system's database compare utilities enable administrators to compare local data stores at a site with data maintained in the central site system. These data reconciliation reports provide additional data integrity verification; and

(6) the system maintains meta data about each data element in a well-formed packet. Meta data, including data types and field size, is used to insure that the data transmitted matches what is expected based on the study design.

The following describes the system architecture showing the relationships between the system components and the system and its environment, and an abstract data model.

The diagrams presented in the proposed solution are intended to depict the system at a high-level. That is, the diagrams show the major system components, but do not depict every function or system feature. Their purpose is to communicate the overall architecture of the system, while each of the subsystems can be implemented using well known techniques in the art.

Introduction to the Proposed Solution

The system will be constructed using a suite of components that comprise most of the system's primary functionality. Much of the functionality created in these components will be directed by parameters stored in local data stores.

Data Capture Component Process Descriptions

1. **Browser**: The rdeBrowser component is the container object for the system. It contains the other components that comprise the system. This component provides a framework that houses and provides access to the rest of the system.

2. **Security Manager**: The security manager component is used to verify a user's access to the system in addition to dictating what rights that user has regarding system resources. In addition, the security manager provides a user interface that enables administrators to add users to the system and assign a user to a group. Administrators can also assign rights to each of the groups supported by the system.

- Data Store:
    - Security Database
- Data In:
    - UserID, Password
    - Security profile from security database
- Data Out:
    - Access admitted or denied
    - Access control list (ACL)

3. **Document Manager**: The document manager controls all data flowing into and out of the forms within the Browser. It also creates the dynamic portions of the document and sets document attributes. This component is used to save and retrieve the clinical data entered into the forms.

- Data Store:
    - EDCData Database
    - Validation database
    - Form Repository

- Data In:
    - Clinical data used with the forms from the EDCData database
    - Configuration information from the configuration manager
    - Validation information from the validation manager
    - Form information loaded from the form repository
- Data Out:
    - Edited or new clinical data records
    - Results of data validation
    - Dynamically updated forms

4. **Data Manager**: The data manager will generate and store data about the clinical data and the processes that support it. For example, the data manager will track when backups/recoveries are performed as well when the system transmits data. Furthermore, the data manager generates tracking IDs and CRCs to ensure that transmitted data arrives correctly at the central site system (CSS).

- Data Store:
    - SysAdmin database
    - Message database
- Data In:
    - RdeData database
    - Transmission status
- Data Out:
    - Data statistics
    - Message status
    - System status

5. **Error Manager**: The error manager will log all system errors, report serious errors to the help desk, provide a descriptive message for the user, and inform the system as to what action must be taken (i.e., continue, shut the system down, exit current routine). Also, the error manager will shut down the system when severe errors occur. The error manager is used by all other system components.

- Data Store:
    - SysAdmin database
- Data In:
    - Error number
    - Calling procedure
- Data Out:
    - Continue flag
    - Message
    - SysAdmin database error log

6. **Configuration Manager**: The configuration manager component manages system upgrades and establishes the choices for the configurable components of the system. The configuration manager uses configuration rules to establish certain system behavior.

- Data Store:
    - SysAdmin database
- Data In:
    - Configuration rules from the SysAdmin database
    - User options for user configurable system features
- Data Out:
    - Set configuration properties

7. **Message Manager (Hermes)**: The message manager will control all data packets sent to and received from the CSS. The message manager works with the communication system to send and received event blocks.

- Data Store:
    - Message repository
    - Message database
- Data In:
    - Messages from the communication system
- Data Out:
    - Formatted messages

8. **Reporting System**: The reporting system will enable users to view and print a variety of standard and custom reports. These reports will show information regarding site data as well as site administrative information.

- Data Store:
    - EDCData database
    - SysAdmin database
    - Message database
- Data In:
    - EDCData data
    - SysAdmin data
    - Message data
    - Record filter data (i.e., date ranges)
- Data Out:
    - Reports

9. **Validation Manager**: The validation manager will use validation rules stored in the validation database to evaluate form data. The validation manager is called from the document manager when the user attempts to save clinical data.

- Data Store:
  - Validation database
- Data In:
  - Form data passed in from the, document manager
  - Validation rules from the validation database
- Data Out:
  - Error numbers
  - Error messages
  - Pass or fail validation indicator

10. **Help System**: The help system provides context-sensitive help to users. This help information is presented to the users when they press F1 or click on the help button. The help information includes system images used to instruct users on how to use the system.

- Data Store:
  - Help file repository
- Data In:
  - Context information
- Data Out:
  - Help file

One embodiment of the computer-readable medium of the present invention, i.e., the MetaTrial program, allows for the following functions/features:

Data Objects

a) Flexible Data Base design by a power user (CDM) • Table definition • Column definition • Domain definition • Unique identifier definition (e.g. pat_number + visit code) • Index definition • Trigger and stored procedure definition
INVESTIGATOR SITE INFORMATION
b) Investigator staff information
c) Support of partial date
d) Support of different field level security • Mandatory "always" (e.g. Patient initials) • Mandatory but with the possibility to leave it empty by clicking the related N/A (not available or missing) check box • Mandatory but with the possibility to leave it empty by introducing the related explanation
e) Support of various international date/time formats
f) Support of various international numbering formats
g) Support of various international measure units
h) Support of automatic conversion of values expressed in different international measure units (e.g. inch->cm)
i) Comment/free text page to be filled by the investigators
j) Comment/free text field to be filled by the CRA • Available at Page (master screen) level • Available at Record (detail screen) level • Available at Field level
k) Comment/free text field to be filled by the MRA • Available at Page (master screen) level • Available at Record (detail screen) level • Available at Field level
l) Audit trail information at field level including : • user coordinates • date/time of modification • old and new value of the amended field • reason of amendment when requested

Data Management

a) Flexible Data entry screens designed by a power user (CDM)
b) Flexible Application menu definition by a power user (CDM) depending of the user profile
c) Automatic patient numbering • by study number + investigator site number + sequential patient number (with the possibility of including existing preset study/center numbers)

d)	Data entry screen sequence customizable to follow a logical decision tree
e)	Remote Data entry Architecture (1 DBMS + Intranet) • Client/Server technology • WEB based technology
f)	Remote Data entry Performance (1 DBMS + Intranet) • Data Entry screen available in 5 [s] max • Data Entry screen data submission in 2 [s] max
g)	Screen to review the Audit Trail data
h)	Data Import/Export • System for exchanging data between databases
i)	Spelling on line
j)	Help on line
CUSTOMIZABLE HELP	
k)	Customizable error/warning messages • Associated at field level (activated by moving from one field to another) • Associated at page level (activated by clicking on save button)
l)	Flexible Error messages definition • Stop message (blocking) • Warning messages (not blocking)
m)	Data consistency checks (at screen level) • Data field validation • Flexible cross-field validation between different screens • Patient duplication checks
n)	Audit log to trace modifications to CRFs
o)	Flexible Record Data entry screen definition • Tabular form • Full screen form
p)	Support of a Drug Medication dictionary, available at Data entry stage on-line, e.g. WHO-DRL multi-language)
q)	Page (screen) navigation • Page by page (sequential next and previous page) • Direct access through a full index page
r)	Index Page showing the CRF page status : • Never entered • Entered but not locked by the CRA • Locked by the CRA for monitoring, review • Locked by the CRA , ready to be queried • Locked by the CRA , queried successfully
s)	Date format Year 2000 compliant
t)	Diary facility with scheduling capacity
u)	Ability to close/freeze access to the data for a given study and to archive it, either at the full study lock, interim analysis or for a subset of data

Data Reporting Tools for Monitoring

a)	Ad hoc reports customizable • by the CDM • by all users (includes user friendly search by example engine)
b)	Flexible high quality full CRF printout definition
c)	Support of American and European paper format
d)	Support of the local format date/number/unit measure
e)	Save/Run report and its query definition
f)	On-line query/report performance • Response time in 20 [s] max
g)	Export report to pdf, Word, Excel
h)	Remote Data query function
i)	Graphic analysis facility

Data Monitoring

a)	Ability for the CRA to electronically inform the site personnel about the details of each problem found at page/field level
b)	Ability to change the query result status when the problem has been solved
c)	Able to record the number of times that a problem has been detected

Data Workflow

a)	Workflow definition and edition • Data Entry users (investigators) • CRA user • MRA user • CDM user • Read only user • Database Administrator
b)	Flexible User profile definition • Data entry screen defined by CDM • Menu tree defined by CDM • Filter definition by user profile at field level • Available Query/Report list user-defined
c)	Tracking of all activities • Alarms • Warnings
d)	Flexible workflow modification

e)	**Transparent access to the database from investigator depending on the user profile (e.g. the Investigator X working for the center Y must have access to the data of his own center only)**

Additional Points of Interest

a)	Data validation • **Definition of an automatic data validation query plan** • **Running of the data validation query plan by a CDM or a CRA** • **Reporting of the query results in a printout** • **Tracking of data query status with the possibility of manual deactivation**
b)	Software validation • **Source code fully validated** • **CRFs designed without modifying source code, therefore eliminating the need for revalidation**
c)	**Integration with SAS**
d)	Adverse event coding • **Integrated solution for Adverse event coding respecting the standard dictionaries (WHOART, MEDDRA)** • **Autocoding system for Adverse event coding respecting the standard dictionaries with text recognition**

Certain functional requirements of an embodiment of the computer-readable medium of the present invention are as follows:

7 Messaging Sub-System Functional Requirements

7.1 Transmission

7.1.1 In general, the transmission system must support a wide variety of message types. The flexibility of the transmission mechanism will facilitate evolving messaging requirements and message types as yet unknown.

7.1.2 The system must be capable of transmitting data to the CSS in a timely, reliable fashion. Furthermore, the system must provide adequate information to verify that the transmissions have been sent and correctly received.

7.1.3 The system must support both real-time and store-and-forward data transmission architectures.

7.1.4 The CSS must send a receipt for each transmission received from a site. The site uses this information to verify that the transmission was received from the CSS.

7.1.5 The system must be able to generate a transmission disk with production data in case the phone line or modem is not available.

7.1.6 The system must provide the site and monitor feedback showing that the data was received by the CSS.

7.1.7 The system must secure the data transmission files prior to transmission.

7.1.8 The system defines transmissions to include any data or information sent from a site to the CSS or to a monitor. Examples of data transmitted from the site to the CSS include:

7.1.8.1 Clinical data

7.1.8.2 Meta data

7.1.8.3 Transmission data

7.1.8.4 Queries

7.1.8.5 Remote locking

7.1.8.6 System updates

7.1.8.7 Database backup files

7.1.8.8 Electronic help desk requests

7.1.8.9 E-mail correspondence

7.1.8.10 Announcements / Newsletter

7.1.8.11 Other message types as yet undefined

7.1.9 The CSS must be capable of transmitting system upgrades to each site. Examples of system upgrades include:

7.1.9.1 Updated forms

7.1.9.2 Procedures to update the database structure

7.1.9.3 Procedures to update the system configuration

7.1.9.4 Site announcements or important information that must be presented to users as they logon to the system.

7.1.10 The system must enable the sites to verify transmissions received from the CSS.

7.1.11 The system must be capable of transmitting data in the background while the user is entering data.

7.1.12 The system must also be capable of transmitting data in batch mode, while no one is logged into the system. The user or the system may schedule these batch transmissions.

7.1.13 The system must permit users to override the transmission mechanism. The system must prompt the user why he/she elected not to transmit data.

7.1.14 The system must provide a low-tech transmission alternative for cases when a system cannot transmit in the normal fashion due to technical difficulties.

7.1.15 The system must be capable of compressing transmission files.

7.1.16 The system must support a number of different transmission protocols. The system must support a variety of file transfer protocols.

7.1.17 The system must support use over the Internet.

7.1.18 The system must encode transmission files in a pre-specified file format. The file type must identify the data in the file.

7.1.19 Each transmission file must contain header information. The file header must include the following:

7.1.19.1 The message type

7.1.19.2 The date the data file was created

7.1.19.3 A CRC and digital signature for the file

7.1.19.4 The length of the file

7.1.19.5 The site that sent the file

7.1.19.6 The study that the file originated from

7.1.19.7 The user that requested the creation of the file

7.1.19.8 The time the file was created

7.1.19.9 The batch identifier for the file

7.1.19.10 The guide or batch ID for the file

7.1.19.11 Indicator signifying if an ACK is requested

7.1.19.12 A retransmission indicator (retransmission count, 0 if this is the original)

7.1.19.13 The version number of the system creating the file

7.1.20 Each transmission file must contain trailer information. The file trailer must include

7.1.20.1 Visual indicator marking the end of the file

7.1.20.2 End of file marker

7.1.21 The system must provide a mechanism for sending test transmissions to the CSS. It is essential that the administrators be able to test the system at the site, especially the transmission system.

7.1.22 The system must enable monitors and administrators to remotely query a remote user site's system. That is, the CSS must be able to send a message that contains a query to be executed at the site. The remote user site must execute the query and return the results to the CSS in a response message. Remote events such as these must be logged at the remote user site.

7.1.23 The system must be able to remotely initiate backups and schedule transmissions. That is, monitors and administrators must be able to use a message to initiate an internal backup or to schedule a transmission.

7.1.24 The system must be able to send sites an executable that the system run. That is, administrators must be able to send a message that contains a small program with instructions of how and when to execute the program. The system at the remote user site must be capable of running the program at the appropriate time and under the appropriate conditions.

7.1.25 The system must enable monitors to review data that does not pass validation. If the monitor decides that the data is acceptable, they must initiate a message that is sent to the remote user site to turn off the not valid data indicators. That is, the data must not remain flagged as exceptional if the monitor has approved the value.

7.1.26 The system must enable inter-site communication. That is, the system must allow users at one remote user site to communicate with users at other remote user sites.

7.2 Transmission Manager

7.2.1 The system will report any transmissions not acknowledged after a threshold (measured in days) has been exceeded.

7.2.2 Optionally, the system must notify the monitor each time one of his/her sites transmits to the CSS.

7.2.3 The system must track the status of all transmissions.

7.2.4 Each transmission file must include a tracking (sequence) number.

7.2.5 The system maintains all transmission files on the system. That is, transmission files are not removed by the system, but are maintained for data redundancy purposes.

7.2.6 The system must periodically transmit data stored in the Transmission Manager to the CSS to ensure that the two systems are synchronized.

7.2.7 The system must record whether or not a transmission was sent after each data entry session. If a user overrides the transmission, the system must record an explanation or reason for the override.

7.2.8 When a user logs onto the system, the system must check to see if the data was transmitted after the last data entry session.

7.2.9 The system must track the status of each transmission. The system must provide a report that lists all transmissions that were not successful or that have not been acknowledged.

7.2.9.1 The transmission tracking report must be available to the data entry users so that they may verify their data transmissions.

7.2.9.2 The transmission tracking report must be sent to the CSS so that the CSS can compare the report to what has been received by the CSS.

7.2.10 The system must be capable of recreating a transmission batch file.

7.2.11 The system must track all the acknowledgements it receives from the CSS.

7.2.12 Negative acknowledgments from the CSS indicate:

7.2.12.1 A garbled transmission file

7.2.12.2 A missing sequence number

7.2.12.3 A transmission file does not match its CRC

7.2.13 The system must synchronize data files between CSS and remote user site each time there is a direct connection between them.

7.2.14 The system must provide a mechanism for browsing the contents of a transmission file.

7.2.15 The system must provide a mechanism for searching the particular text strings within a set of transmission files.

7.2.16 The system must maintain a ranking of transmission data by importance. The system must send the current data at the highest priority. Audit trail information, information about the state of the system, and data warehouse information are all of secondary importance.

7.2.17 The system must utilize dynamic priorities for its messages. That is, the priority of messages will change according to various factors. One such factor is age. The system must support an aging algorithm that increases the priority of messages that are not transmitted after a certain threshold.

7.2.18 The priority system must cause the system to force a large transmission at logon if the remote user site's data has aged beyond a specified threshold.

7.2.19 The priority assignments must be configurable. That is, there may be the need to alter the priority of transmissions at a site or sites due to specific study initiatives.

7.3 Query System for Remote Monitoring

7.3.1 Monitors must be provided with a mechanism to support remote data queries. These queries request clarification or changes to data that does not pass the validation rules. Queries generated at the monitor's office must be transmitted to the site.

7.3.2 Monitor queries must also be logged in the CSS. That is, the CSS must also maintain and audit trail of monitor queries.

7.3.3 Users at the remote user site must also be able to query monitors for information regarding study data or procedures. That is, the query communication mechanism must be two-way.

7.3.4 Each query generated must be assigned a unique ID by the system. This guide is used to manage and track each query.

7.3.5 As users logon to the system, the system must notify them of outstanding queries, transmissions, or messages.

7.3.6 All serious adverse events (Aes) must generate a query to the monitor (see Adverse Events Manager).

7.3.7 The system must generate a report that identifies all forms that contain data that does not pass the validation routines. This report must be sent to the monitor in for form of a query.

7.3.8 The system must permit queries to target the CRF or field level.

7.4 Remote Locking

7.4.1 The system must enable the monitor to initiate a lock from his/her office. That is, the monitor must be able to lock reviewed data remotely.

7.4.2 If a lock is initiated, but the data has changed before the remote user site receives the lock, the lock must not be initiated and a message will be transmitted to the monitor who initiated the lock.

7.4.3 Optionally (through system configuration) the remote user site must have the option of rejecting a lock (assuming its because they have more data to enter or edits to make).

7.4.4 The system must permit the monitors to initiate locks from the remote user site as well as remotely.

7.4.5 A lock must prevent a CRF from being edited.

7.4.6 The monitor must have to ability to unlock data that has been locked. Remote users must not have the ability to unlock previously locked data. Monitors must be able to unlock data remotely or locally.

7.4.7 The system must use a visual cue to notify sites and monitors that a form is locked.

7.4.8 The system must support locking at several different levels. Locking granularity includes:

7.4.8.1 CRF

7.4.8.2 Patient

7.4.8.3 Visit

7.4.8.4 Phase (a sequence of visits)

7.4.8.5 Field

7.4.8.6 Study

7.4.9 The system must enable monitors to mark data as reviewed. This capability must be provided remotely and locally.

7.4.9.1 If a subsequent edit is made the monitor must be notified and the form marked as unreviewed.

7.4.10 The system must provide a visual cue to tell the site staff and monitors that the data has been reviewed.

7.4.11 The system must be able to lock the entire database to facilitate closeout

7.5 Remote Tech Support

7.5.1 The system must provide users with an electronic form to request technical support from the help desk. The form will hold the following data:

7.5.1.1 Site ID

7.5.1.2 Study ID

7.5.1.3 User ID or the requesting party

7.5.1.4 Problem category

7.5.1.5 Problem description

7.5.1.6 Best time to contact by phone

7.5.1.7 Indicator specifying whether the problem is preventing them from entering data or causing the data to be lost or corrupted.

7.5.1.8 Telephone number (pre-filled by the system but can be changed)

7.5.2 The system must maintain a log of all electronic support transactions. That is, the system must maintain an audit trail of all electronic help desk transactions.

7.5.3 The electronic help desk form reply must have the ability to include shortcuts to system utilities.

7.5.4 The electronic help desk support form must enable users to capture screen images to send to the help desk.

7.5.5 The system must provide the ability to update the help system remotely to include new information learned from the help desk.

7.5.6 The electronic technical support system must enable the help desk to periodically send users a FAQ.

7.5.7 Optionally, the electronic help desk feature must CC the monitor on all requests for assistance and help desk responses.

7.5.8 The system must include a tracking number in the help desk responses.

7.6 General E-mail

7.6.1 The system must enable sites to send e-mail. The system will provide pre-built address book that cannot be modified by the users at the site.

7.6.2 The system must enable users to CC appropriate study staff with important messages.

7.6.3 The system must maintain an audit trail of each message sent.

7.6.4 Each message must have a guide for tracking purposes.

The computer-readable medium allows the generation of study parameters, including, but not limited to, forms, data elements, and rules. In addition, the computer-readable medium allows for customization of views of study information, such as, in the case of pharmaceutical clinical trials, study related documents, study management reports, clinical data management reports, drill downs to individual case report forms (CRFs), industry news, study scheduling, etc. In addition, the computer-readable medium, for example, in the case of pharmaceutical clinical trials, enables interactive voice response system (IVRS) capabilities for patient randomization, inventory management, and patient diaries.

The process flow for one embodiment of the computer-readable medium of the present invention, i.e., Hermes, is as follows:

Hermes Process Flow

Startup Process

1. Start Hermes application
2. Hermes already running?
    a. No, continue
    b. Yes, load previous instance and shutdown
3. Process command line options
4. Open transmission database
5. Load communication configuration
    a. Load type of connection (dial-up, LAN, none, etc.)
    b. Load version id
    c. Load encryption id
    d. Load remote access site (RAS) phone book entry
6. Initialize TCP sockets
7. Load Hermes main form
8. Dial-up connection?
    a. Yes

```
            i. Ras available?
                1. No, error
                2. Yes, continue
            ii. Phone book entry available?
                1. No, error
                2. Yes, continue
            iii. Connect using RAS
                1. If error connecting retry RAS_RETRY_THRESHOLD
                   times
            iv. Continue
        b. No, continue
    9. Check communication status, configured to operate on-line?
        a. Attempt to open connection to server, does the server respond?
            i. Yes, the system is on-line
            ii. No, the system is off-line (queues all data for transmission
                when connection becomes available)
        b. No, set system to offline operations
    10. Scan database for data queued for transmission to Central Site System, data
        queued?
        a. Yes, transmit queued data (examine Hermes site transmission process)
        b. No, continue
    11. Scan for data overflow cache, data cached?
        a. Yes, transmit queued data (examine Hermes site transmission process)
        b. No, continue
    12. Initialize a new data transmission buffer
```

Receive Data from Local Browser Process

13. IPC socket gets message from browser

14. Determine message type and process

    a. Request transmission message forces an immediate transmission of any queued data

    b. Logoff message initiates a Hermes shutdown process (see shutdown process)

    c. Freeze message forces the Hermes main form to be frozen (or unavailable to users)

    d. Busy message causes Hermes to report to the browser if it is currently processing information

    e. Ping message causes Hermes to report to the browser whether or not it is in the on-line state

    f. Data messages are cached in a buffer until a transmission is triggered

        i. Log record as received from browser in the database

        ii. If buffer threshold is met the transmit data

Transmit Data to Central Site Process

15. Build data transmission buffer from data record cache

16. Generate a digital signature for the data transmission buffer

17. Build header record

18. Build trailer record

19. Add header and trailer records to the data transmission buffer

20. Encrypt data file?

    a. Yes, apply configured encryption algorithm and continue

    b. No, Continue

21. Initiate timer

22. Transmit data to configured Central Site TCP/IP address

    a. Validate a well-formed packet

    b. Encrypt data (if configured for encryption)

    c. Start transmission timer

    d. Send data

    e. Calculate transmission time, update the session statistics

23. Check return code from Central Site, was transmission successful

    a. Yes, log successful transmission

    b. No, log the error, update the transmission status, retry transmission

Receive Data from Central Site Process

24. Event fires indicating that a message was received from the Central Site system (which is routing the data from other sites)

25. Determine message type?

    a. If message count type, then display the number of incoming messages to expect and set internal counter

    b. If data message the process message

26. Process message

a. Determine encryption type (if any) and decrypt if necessary

b. Check for a well-formed message (message format and trailer)

c. Load the header information

d. Check digital signature and message integrity

   i. If message fails then return failure code

   ii. If message succeeds then return positive return code and cache data record for loading by browser

e. Load message information into transmission tracking database table

System Shutdown Process

27. Check for queued data (including previously failed transmissions)
28. Transmit queued data
29. Transmit disconnect message
30. Close open databases
31. If dial-up connection then hang-up and close connection
32. Close any open forms
33. Close main form and release memory

Central Site Hermes Process

34. Load application and main form
35. Listen for data being transmitted from sites
36. Event fires indicating a data packet was received from a site

   a. Determine message type
   b. If connect message
      i. Add site to connected sites list
      ii. Schedule data transmissions to site
   c. If auto-connect message
      i. Add site to connected sites list
      ii. Schedule data transmissions to site (with less regard for performance impact at site)
   d. If disconnect message
      i. Remove site from connected list
   e. If data message
      i. Determine encryption type (if any) and decrypt if necessary
      ii. Check for a well-formed message (message format and trailer)
      iii. Load the header information
      iv. Check digital signature and message integrity
      v. If message fails then return failure code
      vi. If message succeeds then return positive return code and cache data record for loading by Central Site
      vii. Load message information into transmission tracking database table

37. Event fires indicating a scheduled data transmission (occurs first after a connect message is received from the site)
   a. Check for connected sites where the Central Site has data to transmit to them
   b. For each site with data waiting for transmission

i. Check for valid site id

ii. Check for valid study id

iii. Load the list of data messages waiting to be sent from the database

iv. Send the number of data packets to expect to the site

v. Re-generate the data transmission packet

vi. Transmit the message

    1. Encrypt the message (based on configuration)

    2. Attempt to transmit to site

        a. If attempt fails check error type

        b. If recoverable error, backoff appropriate amount of time and try to retransmit, try SITE_RETRY_THRESHOLD times

vii. Update the transmission status in the database

Although described above with reference to certain specific embodiments, the present invention is nevertheless not intended to be limited to the details shown. Rather, the present invention is directed to a method, system, and computer-readable medium for collecting and processing electronic data and various modifications may be made in the details within the scope and range of equivalents of the claims and without departing from the spirit of the invention.

What is Claimed:

1. A method for collecting and processing electronic data entered at a remote user site comprising:

entering data into a computer at a remote user site using software adapted to receive data entered by a user, process the data, and communicate with a central site system via a computer network;

comparing characteristics of the data to preselected characteristics for each specific type of data to be collected to determine if the data is acceptable;

transmitting an acknowledgement at the remote user site corresponding to the determination of the acceptability of the characteristics;

storing the entered data at the remote user site;

converting the data into packets;

transmitting the data packets to the central site system;

notifying the remote user site of the status of the data transmission; and

storing the transmitted data at the central site system, wherein the steps of the method proceed in real-time and transmission of data does not significantly affect the performance of any step of the method.

2. The method of claim 1 wherein a transmission control protocol/internet protocol is used to transmit the data packets from the remote user site to the central site system.

3. The method of claim 2 wherein the data packets are transferred by hypertext transfer protocol tunneling using port 80.

4. The method of claim 1 wherein each data packet is encrypted before each data packet is transmitted.

5. The method of claim 4 wherein the central site system decrypts the data packets after the packets are transmitted and the central site system identifies the data for grouping with the corresponding fields.

6. The method of claim 1 wherein the data stored at the remote user site is compared to the data stored at the central site system to synchronize the data to determine if data transmission is necessary to synchronize the data and to transmit any data determined to be necessary to synchronize the data between the remote user site and the central site system.

7. The method of claim 1 wherein a network connection between the remote user site and the central site system is intermittent; before the data packets are transmitted to the central site system, the remote user site detects if a network connection is present; and, if a network connection is not present, the remote user site continuously attempts to establish a network connection.

8. The method of claim 1 wherein the data packets are transmitted via a network connection and the computer network is at least one member selected from the group consisting of the Internet, a local area network, a wide area network, an intranet, a dial up connection, and a virtual private network.

9. The method of claim 8 wherein the data packets are transmitted via the Internet.

10. The method of claim 1 wherein the data packets are transmitted continuously while data is being entered into a computer at the remote user site.

11. The method of claim 1 wherein the data is for pharmaceutical clinical trials.

12. The method of claim 1 wherein the data at the central site system is converted into packets, the packets are encrypted, and the encrypted packets are transmitted to remote monitors for evaluating the data.

13. The method of claim 12 wherein the data is decrypted by the remote monitors, the data is evaluated, and the data is altered.

14. The method of claim 13 wherein comments are applied to the data or the data is locked by the remote monitors.

15. The method of claim 14 wherein the altered data is converted into data packets and encrypted, and the encrypted data packets are transmitted to the central site system.

16. The method of claim 15 wherein the central site system decrypts the altered data packets after the packets are transmitted.

17. The method of claim 1 wherein the data packets comprise data elements and separators and the size of each data packet is no more than about 64,000 bytes.

18. A system for collecting and processing electronic data entered at a remote user site comprising:

a remote user site having a computer;

software used on the computer, the software being adapted to enable entry of data by a user, communicate with a central site system via a computer network, compare characteristics for a specific type of data to be collected to specific preselected characteristics to determine if the received data is acceptable, transmit an acknowledgement to the remote user site corresponding to the determination of the acceptability of the characteristics, convert entered data into data packets, and transmit data to the central site system;

a network connection between the computer at the remote user site and a central site system for enabling the software to transmit data from the remote user site to the central site system and to synchronize the data at the remote user site with the data at the central site system;

a storage device at the central site system for storing the transmitted data; and

a storage device at the remote user site computer for storing entered data.

19. The system of claim 18 wherein the software is adapted to use a transmission control protocol/internet protocol to transmit the data packets from the remote user site to the central site system.

20. The system of claim 19 wherein the software is adapted to use hypertext transfer protocol tunneling using port 80 to transfer the data packets.

21. The system of claim 18 wherein the software is adapted to encrypt each data packet before transmission of each data packet.

22. The system of claim 21 wherein the central site system is adapted to decrypt the data packets after the packets are transmitted.

23. The system of claim 18 wherein the software is adapted to compare the data stored at the remote user site with the data stored at the central site system, to determine if data transmission is necessary to synchronize the data, and to transmit any data determined to be necessary to synchronize the data.

24. The system of claim 18 wherein the network connection is intermittent.

25. The system of claim 24 wherein the software is adapted to detect if a network connection is present before the data packets are transmitted to the central site system and, if a network connection is not present, to continuously attempt to establish a network connection.

26. The system of claim 18 wherein the computer network is at least one member selected from the group consisting of the Internet, a local area network, a wide area network, an intranet, a dial up connection, and a virtual private network.

27. The system of claim 26 wherein the computer network is the Internet.

28. The system of claim 18 wherein the software is adapted to transmit the data packets continuously while data is being entered into a computer at a remote user site.

29. The system of claim 18 wherein the data is for pharmaceutical clinical trials.

30. The system of claim 18 further comprising remote monitors for receiving and evaluating data converted into encrypted packets and transmitted by the central site system.

31. The system of claim 30 wherein the remote monitors are adapted to decrypt, evaluate, and alter the data.

32. The system of claim 31 wherein the remote monitors are adapted to apply comments to the data or lock the data.

33. The system of claim 32 wherein the remote monitors are adapted to convert the altered data into data packets, encrypted the data, and transmit the encrypted data packets to the central site system.

34. The system of claim 33 wherein the central site system is adapted to decrypt the altered data packets after the packets are transmitted.

35. The system of claim 18 wherein the data packets comprise data elements and separators and the size of each data packet is no more than about 64,000 bytes.

36. The system of claim 18 wherein the software is adapted to be modified using a computer programming language selected from the group consisting of Visual Basic, C/C++, Java, COBOL, Delphi, VBScript, Java Script, and any language capable of creating ActiveX components of Java applets.

37. The system of claim 18 wherein the network connection has a user authentication device, the device determining if a user has right of access to the connection and communicating to the central site system the right of access determination, the central site system supplying a network connection only to a user determined to have a right of access.

38. The system of claim 18 wherein the storage device at the central site system is a computer hard drive.

39. The system of claim 18 wherein the storage device at the remote user site is a computer hard drive.

40. A computer-readable medium having computer-executable instructions for performing the steps of:

entering data into a computer at a remote user site;

comparing characteristics of the data to preselected characteristics for each specific type of data to be collected to determine if the data is acceptable;

transmitting a notification to the remote user site corresponding to the determination of the acceptability of the characteristics;

storing the entered data at the remote user site;

converting the data into packets;

transmitting the data packets to the central site system via a computer network; and

notifying the remote user site of the status of the data transmission, wherein the steps performed by the computer-readable medium proceed in real-time and transmission of data does not significantly affect the performance of any step.

41. The computer-readable medium of claim 40 wherein, if the remote user site is unable to transmit the data packets to a central site system via a computer network, the computer-readable medium stores entered data at the remote user site.

42. The computer-readable medium of claim 40 wherein the identity of remote users is authenticated based on an access control list.

43. The computer-readable medium of claim 40 wherein the data packets are encrypted prior to being transmitted.

44. The computer-readable medium of claim 40 wherein a transmission control protocol/internet protocol is used to transmit the data packets from the remote user site to the central site system.

45. The computer-readable medium of claim 44 wherein the data packets are transferred by hypertext transfer protocol tunneling using port 80.

46. The computer-readable medium of claim 40 wherein, before transmission of each data packet, each data packet is encrypted.

47. The computer-readable medium of claim 40 wherein the data packets are decrypted after the packets are transmitted and received by the central site system and the data is identified for grouping with the corresponding fields.

48. The computer-readable medium of claim 40 wherein the data stored at the remote user site is compared to the data stored at the central site system to determine if data transmission is necessary to synchronize the data, and to transmit any data determined to be necessary to synchronize the data.

49. The computer-readable medium of claim 40 wherein the network connection is intermittent; before the data packets are transmitted to the central site system, the presence of a network connection is detected; and, if a network connection is not present, the computer-readable medium continuously attempts to establish a network connection.

50. The computer-readable medium of claim 40 wherein the computer network is at least one member selected from the group consisting of the Internet, a local area network, a wide area network, an intranet, a dial up connection, and a virtual private network.

51. The computer-readable medium of claim 50 wherein the data packets are transmitted via the Internet.

52. The computer-readable medium of claim 40 wherein the data packets are transmitted continuously while data is being entered into a computer at a remote user site.

53. The computer-readable medium of claim 40 wherein the data is for pharmaceutical clinical trials.

54. The computer-readable medium of claim 40 wherein the data at the central site system is converted into packets, the packets are encrypted with a digital

signature, and the encrypted packets are transmitted to remote monitors for evaluating the data.

55. The computer-readable medium of claim 54 wherein the data is decrypted, evaluated, and altered.

56. The computer-readable medium of claim 55 wherein comments are applied to the data or the data is locked.

57. The computer-readable medium of claim 56 wherein the altered data is converted into data packets and encrypted with a digital signature and the encrypted data packets are transmitted to the central site system.

58. The computer-readable medium of claim 57 wherein the altered data packets are decrypted after the packets are transmitted.

59. A method for synchronizing electronic data at a remote user site with data at a central site system comprising:

comparing data entered into and stored at a remote user site with data transmitted to and stored at a central site system by transmitting data between the remote user site and the central site system via a network connection;

determining from the transmission if it is necessary to transmit additional data to synchronize the data at the remote user site with the data at the central site system; and

transmitting data between the remote user site and the central site system according to the determination step to synchronize the data at the remote user site with the data at the central site system.

60. The method of claim 59 wherein the data is transmitted using a transmission control protocol/internet protocol.

61. The method claim 60 wherein the network connection is intermittent; before the necessary data is transmitted between the remote user site and the central site system, the remote user site detects if a network connection is present; and, if a

network connection is not present, the remote user site continuously attempts to establish a network connection to transmit the necessary data.

FIG. 1

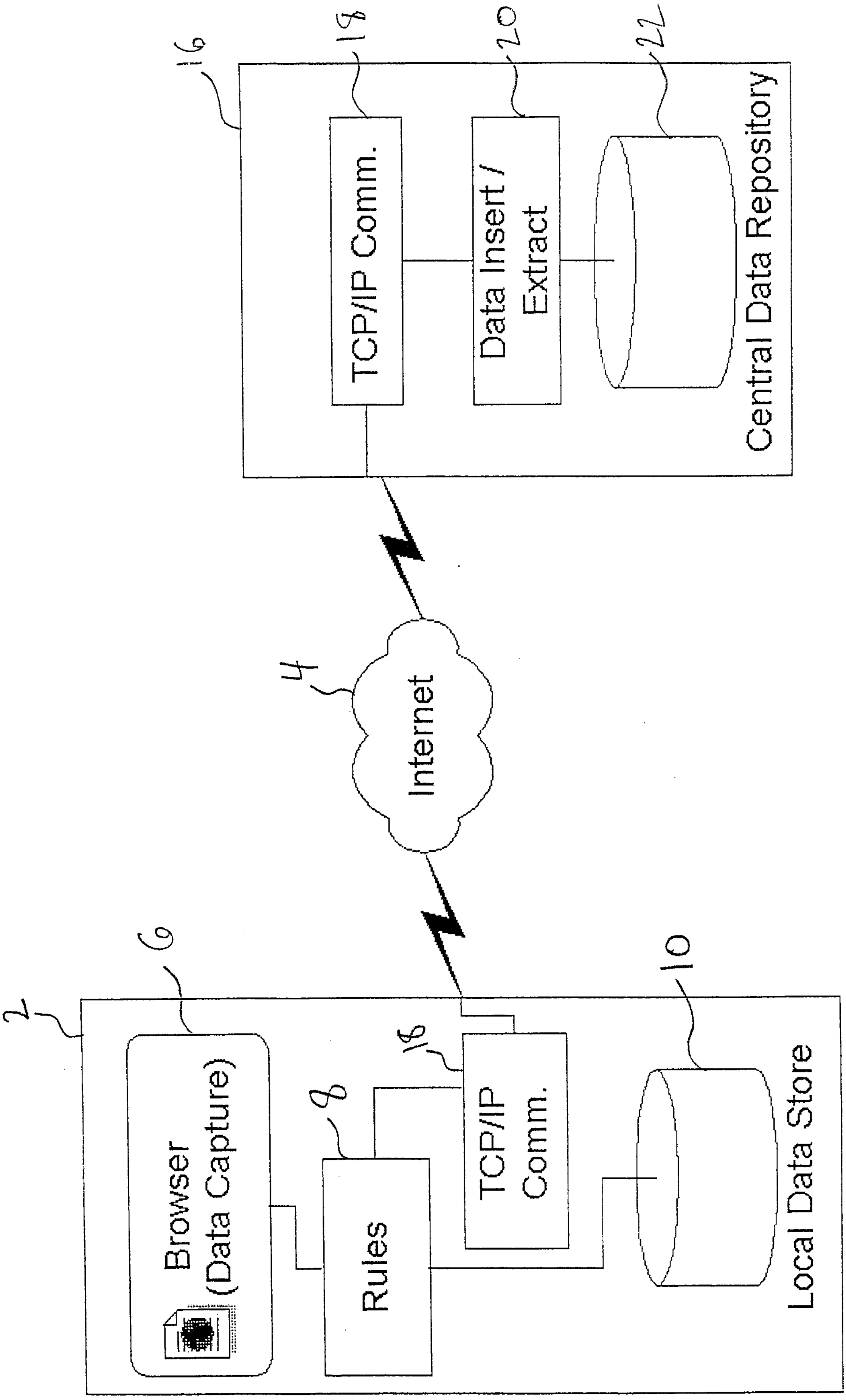
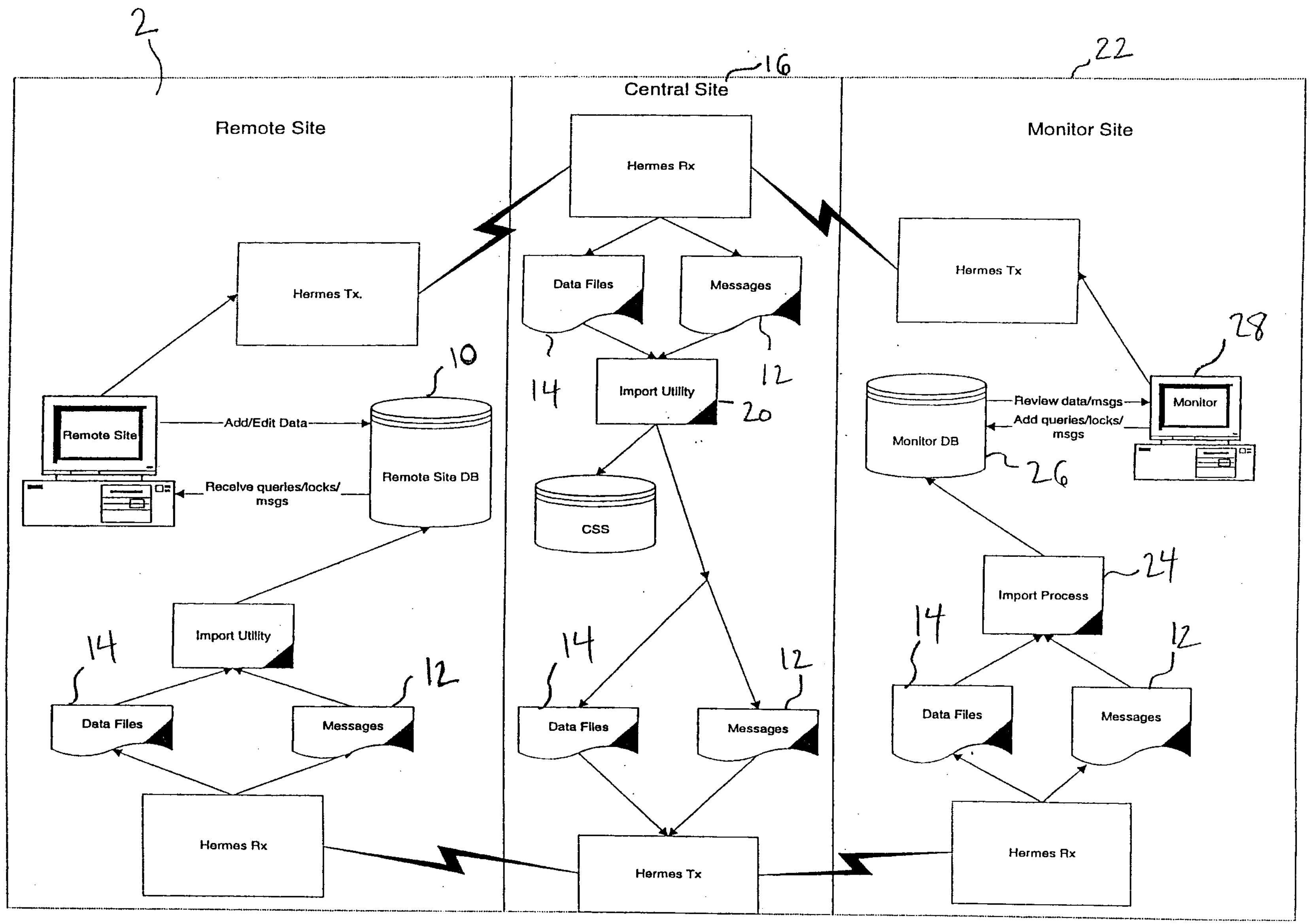

2
6
Browser
(Data Capture)
8
Rules
18
TCP/IP
Comm.
10
Local Data Store
4
Internet
16
18
TCP/IP Comm.
20
Data Insert /
Extract
22
Central Data Repository

Fig. 2


2
16
22
Remote Site
Central Site
Monitor Site
Hermes Rx
Hermes Tx.
Hermes Tx
Data Files
Messages
14
12
Import Utility
20
Remote Site
Add/Edit Data
10
Remote Site DB
Receive queries/locks/msgs
CSS
Monitor DB
26
Review data/msgs
Add queries/locks/msgs
Monitor
28
Import Process
24
Import Utility
Data Files
Messages
14
12
Hermes Rx
Hermes Tx
Hermes Rx
Data Files
Messages
14
12
14
12

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/27020

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) :Please See Extra Sheet.
US CL :Please See Extra Sheet.
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : Please See Extra Sheet.

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

NONE

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 5,742,905 A (PEPE et al) 21 April 1998. abstract, figures 1-3, 21, 24, 28-35, and 45, column 5 line 28 to column 7 line 15, column 21 line 64 to column 22 line 62, column 24 line 14 to column 27 line 12, column 30 lines 27-56, column 34 line 17 to column 35 line 30, and column 36 lines 38-51.	1-61
Y, P	US 6,067,561 A (DILLON) 23 May 2000. abstract, figures 1-2, column 5 line 60 to column 8 line 65, and column 8 line 66 to column 10 line 57.	1-58

☒ Further documents are listed in the continuation of Box C. □ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

01 NOVEMBER 2000

Date of mailing of the international search report

09 JAN 2001

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231
Facsimile No. (703) 305-3230

Authorized officer
AN, MENG-AI T. James R. Matthews
Telephone No. (703) 305-9678

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y, P	US 5,983,350 A (MINEAR et al) 09 November 1999. abstract, figures 2-4, column 5 line 34 to column 6 line 26, column 10 lines 30-60, and column 11 line 46 to column 12 line 48.	4-5, 12-16, 21-22, 30-34, 43, 46-47, 54-58
Y	US 5,647,002 A (BRUNSON) 08 July 1997. abstract, figures 1-2, and column 4 line 45 to column 6 line 37.	6-8, 23-26, 48-50, 59-61
A	US 5,848,415 A (GUCK) 08 December 1998.	1-61
A	US 5,826,023 A (HALL et al) 20 October 1998.	1-61
A	US 5,822,526 A (WASKIEWICZ) 13 October 1998.	1-61
A	US 5,790,790 A (SMITH et al) 04 August 1998.	1-61
A	US 5,675,507 A (BOBO, II) 07 October 1997.	1-61

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/27020

A. CLASSIFICATION OF SUBJECT MATTER:
IPC (7):

G06F 13/00, 13/38, 15/16, 15/17;
H04Q 7/00, 7/20

A. CLASSIFICATION OF SUBJECT MATTER:
US CL :

709/200-207, 217-219, 225, 227-229, 245-246;
707/10, 103;
713/200-201;
455/4.1, 4.2, 412, 418-419, 461-463

B. FIELDS SEARCHED
Minimum documentation searched
Classification System: U.S.

709/200-207, 217-219, 225, 227-229, 245-246;
707/10, 103;
713/200-201;
455/4.1, 4.2, 412, 418-419, 461-463

B. FIELDS SEARCHED
Electronic data bases consulted (Name of data base and where practicable terms used):

WEST SEARCH-- > Search Terms-- > electronic data, remote user, network, data/accept?, acknow ledgement, convert?/data/packet, status/data transmission, protocol, HTTP, packet/encrypt?/decrypt?, synchronize/data, establish/network connection, Internet