APPARATUS AND METHOD FOR INTEGRATING, SORTING AND ANALYSING HETEROGENEOUS DATA SETS

A system allows data stored on existing disparate data systems (1) to (4) to be extracted, standardised and integrated such that different organisations, areas and individuals can be compared against each other. In particular, the system allows heterogeneous data sets to be standardised and transmitted to a processing means (7) which can infer whether the standardised data sets are out with a normal reference criteria or range and can alert a relevant body or person to the occurrence of an adverse event which may be associated with a location, organisation or individual.
Published: without international search report and to be republished upon receipt of that report

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.
Apparatus and Method for Integrating, Sorting and Analysing Heterogeneous Data Sets.

The present invention relates to apparatus and method for extracting data sets from remote, disparate locations in order that the data can be analysed and used to report the occurrence of adverse events. In particular the described apparatus and method can be used as an early warning system to predict where an adverse event may occur.

Clinical care within the NHS is generally of a very high standard. However it would be unrealistic to expect that accidents and mistakes do not occur during the provision of health care in hospitals, surgeries and in the community. There is an inevitable risk that, as medical treatment relies largely on human intervention, mistakes will occur from time to time. Unfortunately, when mistakes are made by employees within the health service the consequences tend to be far more serious than they are in other occupations.

Although the majority of employees within the Health Service have high clinical, ethical and professional
standards, instances of malpractice and incompetence do occasionally occur. The consequences of these events can be both devastating and distressing to the individual patient and their families. This has been particularly apparent over recent years as disturbing cases of clinical malpractice are highlighted by the media.

It is estimated that nearly 28,000 written complaints are made about aspects of clinical treatment in hospitals, and 400 people die or are seriously injured in adverse events involving medical devices every year. The occurrence of such serious failures, not only undermines public confidence, but also costs the Health Service an estimated £400 million a year in settling clinical negligence claims. Furthermore it is unclear whether these official statistics reveal the true scale of these problems. Many incidents remain unreported or unnoticed, and incident reporting systems are particularly poor within primary care. Most incidents which are reported tend to be of a serious nature, where serious malpractice or injury has occurred to the patient and reporting of “near-misses” are almost non-existent. In this manner, any action taken tends to be reactive to incidents which have occurred as opposed to proactively trying to prevent future incidents from occurring.

It is therefore essential that there is a facility is in place to report when and where such incidents take place, in order that action can be taken to prevent any further such incidents. There are at present a number of reporting systems in place within the NHS. These include local and regional reporting schemes, complaints systems, external reviews, incident enquiries and audits. However
there is little consistency in these systems around the
country. It will be appreciated that within the NHS,
there exists a number of different health care providers
including doctors, nurses, community workers and
affiliated occupations, working within a number of
different health care environments such as hospitals,
surgeries and in the community. Consequently there
exists a diversity of data record storage and reporting
systems (Patient Administration Systems) which differ in
format and content.

It would be advantageous to integrate and standardise
these data systems, as this would permit comparisons to
be made regarding the performance of individual providers
and would also allows areas to be identified where there
may be cause for concern. However there is no single
system that can collate this information in order to
assess the incidence of adverse effects around the
country so that different trusts, organisations or
individuals can be compared.

Unfortunately there are various obstacles to the
centralisation or sharing of such data systems. It is a
recognised problem that there is an incomplete IT
structure within the Health Service and there are limited
resources or IT expertise. It will be appreciated that
it would be unduly expensive to provide every provider
with a new system, and that it is likely that there would
be some reluctance to change existing work practices and
embrace complex new systems. There is also the question
of patient confidentiality, and accordingly any data used
in an analytical or reporting manner would have to be
secure.
There is therefore a need for a system which can extract data from existing disparate sources in a secure and confidential manner, in order that such data can be standardised into a format which can be used to predict areas, organisations or individuals where mistakes or adverse events are occurring frequently or repeatedly, in order to warn against clinical problems or malpractice and to prevent such adverse events from taking place in the future.

It is an object of the present invention to provide a system which can intelligently extract data from a plurality of remote disparate locations.

It is a further object of the present invention to provide a system which allows data stored on existing, disparate data systems, to be extracted and integrated such that different organisations, areas and individuals can be compared against each other.

It is a further object of the present invention to provide a system which allows data stored on existing, disparate data systems, to be extracted and integrated such that it can be analysed against set criteria or benchmarks in order to assess the risk of an adverse effect occurring.

According to the present invention, there is provided a data analysis system for use in determining the occurrence of adverse events represented in heterogeneous data sets, the system comprising:
a) means for formatting and transmitting said heterogeneous data sets to a central processing means in a standardised format,
b) inference means for determining whether the data sets are outwith a "normal" reference criteria or range,
c) reporting means for reporting adverse events associated with locations, organisations or individuals.

Preferably the system comprises a centralised platform and one or more local platforms.

Preferably the centralised platform receives data in a standardised format from the one or more local platforms.

Preferably the one or more local platforms act as a gateway through which data must pass in order to be standardised.

Preferably data is transmitted from the one or more local platforms to the centralised platform.

The system may be used to report and predict the occurrence of adverse events within a community, wherein the community is typically an official organisation such as the National Health Service, Civil Service or Police Force.

The central processing means typically comprises at least one database.
The heterogeneous data sets are typically located within a plurality of remote locations.

Preferably the heterogeneous data sets are extracted, formatted and transmitted from the remote locations to the central processing means by one or more acquisition server.

Preferably the heterogeneous data sets are formatted into a standardised format by an acquisition server schema. Typically the one or more acquisition server schema hold a plurality of mapping tables.

Preferably data is standardised within the system by the attachment of codes.

Most preferably the one or more acquisition servers holds national coding standards and mappings of local data.

Preferably the system also comprises a national code sets warehouse.

Preferably the national codes held within the local acquisition server are synchronised with the national codes held within the national code sets warehouse.

Preferably a normalisation engine attaches national codes to the data.

Preferably the system also comprises a data cleanse application. Most preferably the data cleanse application will extract irregular data and send it to a
suspense area. The data can then be accessed for correction or re-submission.

Preferably users can upload data directly to the acquisition server.

The inference means is typically a hardware or software engine which operates in programming language applicable to the computer system or network with which it is used.

The inference means typically applies a series of rules in order to analyse the risk of an adverse event occurring and to predict where an adverse event may occur. For example, the inference means may calculate the "normal" or average number of deaths per year of patients having a particular illness, and then deduce which, if any of the data sets deviate widely from this number.

The inference means may also comprise a statistical analysis programme.

Preferably the central processing means is adapted to receive information from printed source material.

Preferably the criteria and ranges against which the data sets are assessed are easily adjustable.

The data may be communicated from the remote data locations to the central data storage means in an encrypted and secure format. The data may be communicated from the remote data locations to the
central data storage means via an existing database

system, such as NHSnet.

The data may be communicated in XML.

Preferably the reporting means can be escalated.

Preferably the system has an incident acknowledgment and
closure application.

According to a second aspect of the present invention
there is provided a computer program comprising program
instructions which, when loaded into a computer,
constitute the data analysis system according to the
first aspect of the present invention.

According to a third aspect of the present invention
there is provided a method for determining the occurrence
of adverse events from heterogeneous data sets comprising
the steps of:

a) formatting the heterogeneous data sets into a
standardised format;

b) transmitting the standardised data sets to a central
processing means,

c) inferring whether the standardised data sets are
outwith a normal reference criteria or range, and;

d) reporting the occurrence of an adverse event
associated with a location, organisation or
individual.

An example embodiment of the present invention is
described with reference to the following figure:
Figure 1 is a schematic diagram of a simplified exemplary embodiment of the system in accordance with the present invention, and

Figure 2 is a more detailed schematic diagram of the extraction, transfer and processing of data from one remote data location to a relevant body or individual.

Referring firstly to Figure 1, a plurality of remote data locations (1) (2) (3) and (4), data extraction formatting and transmitting means (5) and (6), data processing means (7), and indicating means (8) together form an exemplary embodiment of the present invention. The exemplary embodiment is for use within the National Health Services and the remote data locations (1) to (4) may be clinical based patient administration systems (PAS), which presently exists within hospitals, surgeries and trusts. These existing patient administration systems contain information and data which relate to patient medical history, such as treatment schedules, practitioner’s name and location, and locations where particular events take place. However the format and information contained on these systems will usually vary markedly. In addition the actual systems and IT infrastructure will generally vary from location to location. The system of the present invention can determine the occurrence or incidence of adverse events or “clinical incidents” within the Health Service. Specifically the system extracts, standardises and transmits heterogeneous data sets thereby allowing the data sets to be analysed against set criteria, enabling clinical incidents to be identified, and most importantly reported to a relevant
party to ensure that action is taken to remedy the
incident, and prevent further incidents from taking
place.

The system of the present invention is divided into two
main systems, a National Information Delivery Platform
and Local Data Acquisition Platforms. The National
Information Delivery Platform is a centralised
information centre into which normalised, standardised
and validated clinical incidents are loaded, and this
system will host all the data against which the present
invention analysis and reports will process.

The Local Data Acquisition Platforms (5) are located in
various DH locations and act as a gateway through which
incident data must pass. All incident data generated
will be validated, normalised and standardised through
these systems. Only when these processes have been
achieved will the incident data be allowed onto the
National Delivery Platform.

The data extraction and formatting means extracts
predetermined data sets from the main types of patient
administration systems (1) to (4) automatically. It is
envisaged in the present invention that it would be
advantageous if the type of data which is extracted could
be adjusted to suit the needs of the relevant body or
individual (10) to which the warning would be given. For
example where the system is used within the national
health service, it is envisaged that the Department of
health and/or NHS Executive would have an active role in
deciding the nature of the data which is to be reported
on. The system also is also adapted to receive paper
returns (9). Once the information is extracted either
together from the existing administration systems or
via paper returns, it is standardised by translation
means (5) and transmitted via an input means (6) to the
central processing means (7). It will be appreciated
that in alternative embodiments the translation means (5)
and input means (6) may be combined.

Typically within the system an acquisition server schema
will be built to hold a series of mapping tables which
will be used to maintain a universal keying system
throughout the entire system deployment. Because
incident data will be uploaded from a large variety of
sources and locations (1) to (4), it is important that
the data is normalised into a common format before it is
entered into the National Delivery system, which in the
embodiment is represented by (7). The Acquisition Server
Schema will not only hold the national coding standards
relevant to the National Delivery System, for example
readcodes, OCS, ICD and/or OPCS, it will also contain
mappings of local data (11) relevant to the Local Data
Acquisition Platforms. This is important, for example,
within a hospital where there may be discrepancies in
patient information, i.e. a single patient may have more
than one code across various systems or various
departments. If this is found, the codes will be mapped
to one of the system identifiers for the patient, and it
is that identifier which will be passed with the incident
data to the national database.

It is strongly preferred that the data is transferred in
a secure and encrypted format to prevent tampering with
the system and to ensure patient confidentiality. In one
1 embodiment the data is transmitted via the existing NHS
2 web-server, NHSnet, to ensure maximum security. The data
3 is typically transferred using XML (extensible mark-up
4 language). A document type definition will define the
5 data upload format for interfacing systems, both on and
6 off line. Any clinical incident that is processed by the
7 acquisition servers will be in an XML format conforming
8 to Data Upload DTD.
9
10 The system typically also comprises a national coding
11 synchronisation manager, which synchronises the image of
12 the national codes held within the local acquisition
13 server (5) with the national code sets warehouse (13)
14 that holds the latest versions of the codes. The
15 Synchronisation manager may also be able to upload
16 national code sets from off line storage media.
17
18 The existence of a normalisation engine relating to the
19 Acquisition Server is the key to the success and quality
20 of the information delivered from the national
21 information delivery platform. The engine takes each
22 reported incident and breaks it down into all of its
23 elements. It looks at each of the elements individually
24 and ensures that the relevant national codes are attached
25 to it. For example, an incident may have the following
26 attributes: patient(s), doctor(s), nurse(s), location,
27 ward, diagnosis, drug(s) and disease. The normalisation
28 engine will attach a national code to each of these
29 attributes.
30
31 In the case of a patient, a unique code will be generated
32 if the patient has not already had an incident recorded
33 on the acquisition server. This code will be unique to
the system of the present invention. The patient mapping
table on the acquisition server will be updated with a
reference of the patient's originating code, i.e. the
code held on the originating hospital system. This
originating code will point to the unique code of the
system of the present invention. By maintaining this
mapping table on the acquisition server, other incidents
originating from other systems within the same hospital
can be mapped to the same unique patient code of the
system of the present invention. In the case of the other
attributes, the normalisation engine will find the
appropriate codes from the local image on the acquisition
server of the national codes.

The system also comprises a mapping tables manager which
is used to maintain the local mapping tables on the
acquisition server (5). The mapping tables will be
updated by the mapping tables manager application, the
normalisation engine and by an on-line data cleanse
application (14). The latter is an on-line user
application that is used to clean data that has failed
the validation and normalisation process.

When data has been uploaded from systems that cannot be
reconciled to codes either in the national codes set or
the mapping tables of the system of the present
invention, the normalisation engine will drop these
upload items into a suspense area (15). Users of the
originating system will then be notified by the
acquisition server (5) that there are items in suspense
outstanding. The users will subsequently come on-line
with the data cleanse application and correct the data.
It will be appreciated that in existing systems within the National Health Service it is difficult, if not impossible, to monitor the information entered into existing systems. For example, in a system where it is necessary for a nurse or doctor to enter their unique identification number or a unique patient number, one pre-entered number may be used many times by many different doctors or nurses for the sake of expediency. In the event that the system of the present invention detects that irregular data has been entered, this data will be pulled out and sent back to the transmitting means for correction or re-submission. This can be seen as 14 in Figure 2, and is an essential part of the system of the present invention as it greatly increases data quality within the system.

An on-line upload application is also provided which allows users to upload adverse clinical incident data directly to the acquisition database (5). The normalised data gateway acts as a “pipe” through which the processed clinical incident data will pushed into the national database of the present invention. This incident data will then be deleted from the acquisition database.

The central data processing means (7) decrypts and stores the data. The information is then processed at an initial level by inference in the data processing means (7), which is typically a hardware or software engine, and which applies a set of rules to the data sets in order to predict locations, organisations or individuals where they may be cause for concern. For example, where the system is used within the National Health Service, the inference means may determine the “normal” number of
deaths within a particular group of patients, and then
use this data to determine whether any particular
practitioner or location appear to deviate from this
norm. Alternatively the system may be used to compare
the number of complaints made against different
organisations, locations and individuals. The data may
then be drawn into a statistical analysis component
program to further refine and analyse the data.

The data processing means (7) comprises a delivery
database schema which will hold on a national level all
of the adverse clinical incidents reported through the
various localised acquisition servers. The foundation
for the schema design was driven by the need to create an
analytical platform upon which many algorithms could run.
It is tuned to delivery a variety of results for
reporting a range of formats, including graphical images.
It is also designed to perform a high level of
performance and be extremely scaleable to accommodate
large volumes of data.

The incidents alerts engine is a stand-alone system that
is highly scaleable in that multi instances of the engine
can run on separate machines, upgrading processor power
limitlessly. The engine runs analysis algorithms on the
database, then cross-references the user profiles for
notification parameters that it then sends to the
notification gateway. The alerts engine runs on a
hierarchical model, whereby it determines whether
previous alerts have been acknowledged and where
applicable it will escalate the alert up the chain of
command.
The reporting module (8) will report the results to the relevant body or individual (10). For example, this may be the Chief executive of the particular Trust in which a particular area of concern has been identified. The reporting module (8) may be configured to operate in a number of ways, for example if we have a number of indicator means to highlight possible areas, organisations or individuals of concern. The indicator means may be, in one example, warning lights of red, amber and green, which warn an individual or organisation that action needs to be taken. Alternatively, alerts may be given by e-mail or SMS. For example the reporting means (8) will warn if the results obtained from analysis of a particular data set deviates from a "normal" range or benchmark criteria. The reporting module (8) will issue a warning, for example, to the Chief Executive of the particular trust (10) to monitor the given situation and to take pre-emptive action to remedy the area concerned before a crisis emerges. In one embodiment, the indicator means providing an early warning would be an illumination of a particular colour of light, for example, the amber light. If no action was taken, the warning would progress to red, then the warning would progress to illumination of the red light to indicate that determined action should be mandated.

This "escalating" alert system is seen as a particularly important aspect of the present invention. In one embodiment, the escalation may arise from the use of indicator means as described above. However, in an alternative embodiment, the escalation system may provide warnings in an escalated manner through a hierarchy of organisations or individuals. For example, in the case
of the National Health Service, in the first instance, a Ward Sister may be given a warning and if no action is taken on this warning, a subsequent warning may be given to the Head of the Department in which the Ward Sister works. In this vein, later escalated warnings could be given, for example, to a Medical Director or to the Health Service Head.

A further important aspect of the present invention is the provision of an incidents acknowledge and closure application, which is web based and built with the latest application server technology. High scaleable, robust and performant, it takes advantage of the latest multi-tier platform architecture to provide rapid development environment for building maintainable solutions. This application will be very secure, but at the same time available to browser based clients running in any remote location. Incidents will be updated from the upload state into an acknowledged state during which investigations will be carried out. The incident will next be closed using the same application when the resulting outcome is recorded. In other words, repeated updates will be sent until acknowledgement is obtained from he organisation or individual to which a warning has been sent, that the warning has in fact been received and action will be taken. Closure will only come about once a suitable action has been carried out.

Delivery of the data is carried out on information delivery mechanisms which are built to the last Java Open standards. Developing the Internet applications to J2EE, Suns Java specification ensures the build will be compliant with the leading application and web server
platforms. This ensures that the build will be compliant with new developments in the information delivery arena.

The multiple tier architecture enables the use of application servers to host Enterprise Java Beans that contain the delivery system business logic of the present invention. Keeping this logic separate means the application code can be maintained and upgrade separately from the presentation of that logic. This makes it very easy to accommodate new requirements with a minimum of disruption to a live system. The presentation logic is contained within web servers servlets and JSPs. There is also a separate tier on the architecture that can be maintained separately with all the benefits that this brings. Multiple user connections are managed with the web servers where sessions are tracked and kept secure. Having a multiple tier architecture also ensures multiple options for scaling the system to accommodate very large volumes of users.

It will also be appreciated that particularly in fields such as the Health Service, it is important that all data is transmitted and obtained in a secure environment. The system of the present invention has Internet portal security which is guaranteed with a combination of features included. Netscape servers LDAP authentication system provides secure log-on interface to the portal. This combined with a firewall surround provides a very secure environment to host the national information delivery platform.

An advantage of the present invention is that it overcomes the short-comings of existing systems without
being unduly expensive, as it utilises existing data systems which already exist, but which cannot be used for comparative analysis as they exist in disparate format and content.

A further advantage lies in the fact that the individual health care providers do not need to change their existing recording or reporting systems as the invention provides means for integrating the existing disparate data.

A yet further advantage of the present invention lies in the fact that the system provides a unified and standardised system for reporting and analysing adverse events, using data which already exists in clinically based patient administration systems. The system integrates agreed data sets and allows reporting protocols to be standardised so that an early warning can be given in relation to individuals or areas of concern.

It is recognised that although the present invention has been described in relation to the National Health Service, the system would have useful application in other organisations, for example, within the Police Force or Civil Service.

Further modifications and improvements will be incorporated without departing from the scope of the invention herein intended.
CLAIMS

1. A data analysis system for use in determining the occurrence of adverse events represented in heterogeneous data sets, the system comprising means for formatting and transmitting said heterogeneous data sets to a central processing means in a standardised format, inference means for determining whether the data sets are outwith the "normal" reference criteria or range, and reporting means for reporting adverse events associated with locations, organisations or individuals.

2. A data analysis system as claimed in Claim 1, comprising a centralised platform and one or more local platforms.

3. A data analysis system as claimed in Claim 2, wherein the centralised platform receives data in a standardised format from the one or more local platforms.

4. A data analysis system as claimed in Claims 2 – 3, wherein the one or more local platforms act as gateway through which data must pass in order to be standardised.

5. A data analysis system as claimed in Claims 2 – 4, wherein data is transmitted from the one or more local platforms to the centralised platform.

6. A data analysis system, as claimed in any one of the preceding Claims, used to report and predict the
occurrence of adverse events within a community, wherein the community is typically an official organisation, such as the National Health Service, Civil Service or Police Force.

7. A data analysis system as claimed in any one of the preceding Claims, wherein the central processing means comprises at least one database.

8. A data analysis system as claimed in any one of the preceding Claims, wherein the heterogeneous data set are located within a plurality of remote locations.

9. A data analysis system as claimed in any one of the preceding Claims, wherein data is standardised by the attachment of codes.

10. A data analysis system as claimed in any one of the preceding Claims, wherein the heterogeneous data sets are extracted, formatted and transmitted from the remote locations to the central processing means by one or more acquisition server.

11. A data analysis system as claimed in Claim 10, wherein the one or more acquisition servers holds a plurality of mapping tables.

12. A data analysis system as claimed in Claims 10 – 11, wherein the one or more acquisition server holds national coding standards and mappings of local data.
13. A data analysis system as claimed in Claims 9 - 12, comprising a national code set warehouse.

14. A data analysis system as claimed in Claims 9 - 13, wherein a normalisation engine attaches national codes to the data.

15. A data analysis system as claimed in any one of the preceding Claims, having a data cleanse application.

16. A data analysis system as claimed in Claim 10, wherein users can upload data directly to the one or more acquisition servers.

17. A data analysis system as claimed in any one of the preceding Claims, wherein the inference means is a hardware or software engine which operates in a programming language applicable to the computer system and network which it is used.

18. A data analysis system as claimed in any one of the preceding Claims, wherein the inference means applies a series of rules in order to analyse the risk of an adverse event occurring, and to predict where an adverse event may occur.

19. A data analysis system as claimed in any one of the preceding Claims, wherein the inference means comprises a statistical analysis program.

20. A data analysis system as claimed in any one of the preceding Claims, wherein the central processing
means is adapted to receive information from printed
source material.

21. A data analysis system as claimed in any one of the
preceding Claims, wherein the criteria and range
against which the data sets are assessed are easily
adjustable.

22. A data analysis system as claimed in any one of the
preceding Claims, wherein the data is communicated
from the remote data locations to the central data
storage means in an encrypted and secure format.

23. A data analysis system as claimed in any one of the
preceding Claims, wherein the data is communicated
from the remote data locations to the central data
storage means via an existing database system, such
as NHSnet.

24. A data analysis system as claimed in any one of the
preceding Claims, wherein the data is communicated
in XML.

25. A data analysis system as claimed in any one of the
preceding Claims, wherein the reporting means can be
escalating.

26. A data analysis system as claimed in any one of the
preceding Claims, wherein the system has an incident
acknowledgement and closure application.
27. A computer program comprising program instructions which, when lodged into a computer, constitute the data analysis system as claimed in Claims 1 - 25.

28. A method for determining the occurrence of adverse events from heterogeneous data sets comprising the steps of formatting the heterogeneous data sets into a standardised format, transmitting the standardised data sets to a central processing means, inferring whether the standardised data sets are outwith the normal reference criteria or range, and reporting the occurrence of an adverse event associated with a location, organisation or individual.