International Journal of Medical Informatics (2004) **73**, 285—290

ELSEVIER

International Journal of Medical Informatics

www.intl.elsevierhealth.com/journals/ijmi

Security issues arising in establishing a regional health information infrastructure

Roderick Neame[a,*], **Michael J. Olson**[b]

[a] *Health Information Consulting Ltd., Homestall House, Homestall Lane, Faversham, Kent ME13 8UT, UK*
[b] *Taskcare Ltd., Homestall House, Homestall Lane, Faversham, Kent ME13 8UT, UK*

KEYWORDS
Security issues;
Health information infrastructure;
ID numbers

Summary A regional health information infrastructure is being developed in an internally self-governing country which is a dependent territory of the British Crown, is not part of the United Kingdom but is a member of the British Commonwealth. This country of about 70,000 inhabitants (and significant numbers of visitors) within the British Isles shares many functions with the United Kingdom—from the perspective of this paper the key shared functions relate to the infrastructure of the departments of social security, social services, central registry, all health care services and national insurance systems. Although it remains independent in various other respects, for the most part it endeavours to achieve an harmonious legislative relationship with the UK, and with the EU.

One primary goal of the information infrastructure development project is to provide links between community, primary and secondary healthcare services and thereby to ensure integrity of information as it refers to each individual receiving care services. A second goal has been to integrate this environment with various other government functions including the issuing and checking of NHS ID numbers and of national insurance ID numbers, the payment of social welfare benefits, and perhaps with other functions where access to a common list of names and addresses is a significant factor.

This paper outlines some of the issues that have arisen in endeavouring to meet the often conflicting wishes and needs of different groups as regards a health information infrastructure within a general public sector information service.

1. Introduction

The government of this Island country has embarked on a Strategic Information Project (SIP) as a means of supporting the delivery of 'joined up care' to healthcare service users and to provide an underpinning integrity for the health information services. There are some 13 primary care service provider groups, all of whom have a basic computer system from five different original suppliers. Only a few of these primary care systems have been used effectively. There is a single general hospital in the region with a basically functional but poorly integrated information system, as well as a small outpost hospital. Patients in certain categories and sub-specialty areas are not treated on-Island but are transferred to mainland UK hospitals for diagnosis and care. Recently, a new hospital has been built and is being commissioned.

The Island has an autonomous Department of Health and Social Security (DHSS) which is linked with and shares essentially the same structures as the UK DHSS and National Health Services (NHS). It uses (and issues) NHS unique patient numbers and

*Corresponding author.
E-mail address: roddyneame@taskcare.com (R. Neame).

1386-5056/$ — see front matter © 2003 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.ijmedinf.2003.11.010

NHS unique clinician identifiers. All care providers and services record their patients and the care they provide using a computerized system, but each uses their own unique way of identifying individuals (in theory care services should be linked to the formal NHS number of the patient, but in practice this rarely happens—indeed some systems are unable to hold both the NHS number and the internal system identification number. The Island has a social welfare and benefits system that is essentially the same as for the UK. These health and social security services are paid for in a similar way through a national insurance system, which makes use of its own system of unique identifiers.

This network of services makes use of numerous different databases and personal identifier systems. The main obstacles to a 'seamless' health information environment are that:

- Each individual is identified in different ways by the various systems.
- Each system records a different core dataset about the individual.
- When updates are made on one system (e.g. name, address, residency status or to some other parameter (e.g. a death)), they are not linked to advise other associated systems of this update.
- Achieving basic demographic data consistency (e.g. of current address) across all the health care related systems may take months if not years. This does not lend support to the development of integrated care services.
- Gathering of public health data is at best cumbersome and patchy, and at worst absent, so seriously impeding logical planning of services and expenditures.
- Electronic sharing of clinical data is uncertain and risky since:
  - in the absence of any shared table of unique identifiers it is difficult to be sure of the person to whom the data relates;
  - in the absence of a table of professional-patient relationships, and of the authorizations of each of those professionals to access data about the patient, data may be disclosed in breach of patient confidentiality.

This is the situation that is being addressed by the Strategic Information Project.

2. SIP system overview

The SIP provides a framework that is independent of the specific end-user systems which interface with it—in other words it can readily be configured to function with any type of point-of-care, clinical or administrative system. In outline the SIP comprises the following main elements.

2.1. Unique person index

The SIP has implemented a unique person index (UPI) in which each individual using health and social security services is represented. The index lists all active health care users who are in receipt of care services from community, primary and secondary care as well as all those in receipt of social security benefits. The UPI is designed to be the primary database of information about all clients of the Island Department of Health and Social Security. It records the individual's current name and address (with previous and alternate names, and past addresses), as well as unique health service number, unique national insurance number (social security), and unique hospital patient administration system identifier number. There is a 'known to' list which identifies the list of care service providers with whom the patient has a care relationship.

2.2. Authorised user index

The SIP has implemented an authorized user index (AUI) which is designed to be the primary database for determining authorizations and permissions for access to shared information with the SIP framework. The AUI stores basic care service provider demographics, roles they fulfil and links to organizations/institutions they work with, professional qualifications and specializations, national provider code(s) and hospital approved user code(s).

2.3. Laboratory results viewer

The SIP has implemented a laboratory results viewer (LRV) which enables an appropriately authorised user to display investigation results on screen as soon as those results have been approved for distribution by the respective laboratory services.

2.4. Events and encounters index

The SIP has implemented an events and encounters index (EEI) which constitutes a table of the authorized users and care service providers which have a relationship with each person on the UPI. Each event uniquely links a patient with a provider/organisation, and includes a (start and end) date for the encounter.

2.5. Encounter reports datasets

In association with the EEI, the SIP will shortly be implementing an initial set of encounter reports datasets (ERD), whereby pre-specified data sets are captured and reported in respect of specific types of care encounters. These comprise a community health status and services monitoring system that is not at present readily available. Initially the ERDs will relate to hospital inpatient and outpatient events, cancer, births and deaths, but ERDs for other event types, including primary and community care, can be added in exactly the same way. Specific ERDs are planned or under development for immunizations, screenings, health status warnings and alerts, which would be linked to the UPI record for each individual and will contain key clinical parameters regarding that individuals health status (UPI-associated data set): these would be viewable by any care professional providing services to that person.

This system will provide information directly to certain existing or planned future registry functions, such as for births, deaths, immunizations, preventive care, mental health, communicable diseases and so on—the list of registers can be expanded.

2.6. Information integration framework

The information integration framework (IIF) is the 'hub' whereby authorised users can access the system, through a browser-based logon, and with encryption applied to all transactions handled through the framework. The IIF permits selective access to UPI, AUI, LRV, as well as to various documentary resources (reports, alerts, bulletins, advisories, summaries, statistics, education and training materials, etc.). The IIF offers a means of accessing electronic versions of forms (referral forms and datasets, questionnaires, etc.) which may be required for various service-related purposes. And these may be completed (with automatic insertion of patient and provider details) and submitted online where the relevant service is appropriate set-up to receive and process them in this way.

2.7. Updates

The system is shortly to implement a bi-directional update capability, whereby when any demographic parameters (e.g. name, address) on any of the connected systems are updated and in turn update the UPI, the UPI will then offer these updates to all other systems on which that person is represented.

There are numerous important goals that such a system can serve. It can fulfil the needs of patients to have access to seamless public services and it can fulfil the needs of government for effective monitoring of need for and delivery of services. There can be no argument that these functions are vitally important. But at the same time the system could be used in other ways, and these may potentially infringe upon civil liberties. The vital analysis is whether the public and individual benefits accruing from the use of such a system are sufficient to outweigh any actual or potential disadvantages, and how those potential disadvantages can be effectively minimised.

From a technical security perspective, the SIP system has a unique login for every user, and all transactions within the system are encrypted, measures that can readily be enhanced as required, but are seen as adequate and proportionate to guard against the risk of unauthorized access at the present time. The greatest threat to personal information within SIP comes not from unauthorised users of the system, but from inappropriate use by authorized users.

3. Personal privacy issues

The goal of this paper is to explore the implications for personal information privacy and issues of controls over access to personal information—in other words the parameters that surround the information access controls many of which are being embedded in the AUI. The issues of technical security will not be explored further at this time.

3.1. Existence of an individual

The system records names of individuals who are clients of the health and social security sectors and who are contributors to national insurance. Few would argue that the existence of an individual is a privacy issue, particularly when they are consumers of public services and therefore funds. The concept of a unique 'national identifier' for public sector purposes is not unfamiliar and seems consistent with the implicit relationship between individual and State—as long as the identifier is no more than simply an identifier. In the health sector such unique identification is essential to ensure that key information (e.g. test results, shared care data, etc.) is tagged definitively to the correct individual—mistaken identity could prove disastrous.

However issues of concern may arise when several (or all) public databases are indexed with the same common 'key', and related in such a way that

links can readily be established between them for comparing patterns across sectors. This system does create a link between national insurance, health care and social security, but the data that can be viewed by each of these is restricted to name and address only—see below for further discussion.

3.2. Personal demographics

For each individual on the UPI various personal demographics are recorded, primarily, date of birth, alternate names (also known as) and address (plus previous addresses). Updates of these parameters that are undertaken on linked systems will be passed automatically to all other systems with which that individual is known to have a relationship. The benefits of this are self-evident and there are clearly sound reasons for ensuring that if an individual is represented in more than one way, all those apparently separate 'personas' are linked together at the uppermost level for analysis and statistics as well as for integrity—but there may be concerns too.

An individual can identify themselves in whatever way they choose (alias or also known as) and could choose to be one persona for one service, but to 'hide' this from another by assuming a different identity—as long as the purpose was not to defraud or to act illegally. For reasons of integrity the system as presently configured does not permit this (although it could), and all 'alternate personas' are linked to a single identifier. In just the same way an individual may for whatever reason prefer to have one address for some services, but not for others—again this would not readily be accommodated under the present configuration of this system (although again with a minor modification it could be). Where individuals may have valid (legal) reasons for adopting alternate personas, the SIP system (as presently configured) does not support this since all alternate names are accessible to users, and only a single address and date of birth can be maintained at any one time.

3.3. Clinical information

The care provider keeps the definitive record of care services provided on their 'own' system. The SIP system enables synoptic abstracts of those care encounters to be shared for different reasons—for shared care, for reporting and statistics, for maintenance of official records and registers, etc. Each of these requires careful consideration.

The issues raised below will all be exacerbated if clerical, secretarial and administrative staff are able to access sensitive data (as they do in many clinical environments at present) on behalf of the care provider. Further if the data is 'of value' to someone outside the system (e.g. an insurer or employer) a 'black' operation to provide it (for reward) will emerge from within to fill the need.

3.3.1. UPI-associated data set

Present plans look towards the provision of a set of personal clinical data to all those caring for the patient—for example, about preventive care (immunizations, screenings), and warnings/risks. Whilst of undoubted benefit to both patients and care providers, patients might not necessarily wish to share these data. Although not within present plans, the concept of 'warnings and risks' could perhaps in the future be extended to include the predictions from DNA analyses. Access to these data would be limited to those having a duty of care relationship to the patient.

Two possible scenarios are raised for consideration.

1. The individual requires a medical certificate to be completed, for example, relating to employment, insurance, assurance, etc. This would now be 'informed' by data gathered for other unrelated purposes—which it should not. Predictive data (e.g. from DNA testing) would always be subject to statistical interpretation of applicability. The consequence could bring significant disadvantage to some, even where the data (e.g. results of HIV testing) was negative or uncertain. There is a general rule that once data is known it cannot become unknown to an individual—in other words the individual cannot be expected to compartmentalize their knowledge of a person based on the context within which they come to know of specific data items.
2. A patient attends a provider not previously known to them—for example, in an emergency. The care provider cannot access the UPI-associated emergency care information—with possible outcome of a preventable misadventure. Therefore, to be useful the data must be available to all potential providers of care—so bringing the potential risk of informing those (many) who have no ''need-to-know'', and some of whom may have personal and social relationships with the individual.

3.3.2. Data for shared care

Some of the same general considerations apply to shared care information as outlined above. Certain elements of the record of care may be sensitive to

the patient—for whatever reasons. Therefore, the patient should determine what may be shared, and with whom. However, it may not be possible to know with whom that data will be shared, since it refers to future events—and in many cases the provider seeing the patient (e.g. for a referral) may be any member of that care team (e.g. the duty physician). Therefore, the matrix of permissions must either be created for all possible care providers (clearly impracticable) or for general classes of provider (e.g. by specialty, employment, group membership, etc.).

Sensitivities may also exist on the provider side. For example, an issue has arisen over viewing of lab results, and who may view them. The general rule is that all lab results should be viewable by all providers caring for that patient—or by the members of their clinical 'group' (e.g. partners in the provision of that service). However, some physicians are concerned about this and would not wish their colleagues to be able to 'audit' the care they are providing using this tool, a concern that seems to be more prevalent within certain hospital departments. As the system evolves, so these concerns may emerge more widely and relating to more issues.

3.3.3. Data for analysis and registers

The gathering of data, whether for audit of care, review of workforce activity or preparation of community health statistics, is essential to the proper management of the business. However, there are strong arguments for these data to be 'anonymised'—in other words for all personal identifiers to be replaced such that the analyses remain informative and valid but the individual identities are concealed.

Statistics are intrinsically not about individuals, but about communities. As such there is no purpose in including personal identifiers, other than where they support the inclusion of these data within one sub-section of the community (e.g. age range, gender or ethnic group, geographical location, professional group, etc.). However if all the above are included with the data for analysis, it is inevitable that the individual identity will be revealed. The counter argument, that individual identities may be necessary for audit to be effective, can be accommodated by providing a secure key to identify the subject of a record if there is a pressing reason to do so.

Personalized data is clearly essential for maintenance of registers. However unless the registers are required by statute or other legal instrument, individuals should be able to choose whether or not their identity should be added to the register, and therefore whether or not the relevant ERD should be generated.

4. Discussion

There are strong arguments in support of the functions that this system provides both for the individual and the community. But it also raises serious considerations for personal choice and privacy. The essence of personal privacy is that the individual should be empowered to control who knows what about them, and the system described above has the potential to infringe personal privacy unless adequate controls can be agreed and implemented, or unless there are adequate mechanisms for individuals to 'opt out'.

As far as access to personal clinical data is concerned, much of the potential for abuse could be resolved by the implementation of one step whereby the data subject takes control over their own data, and thereby can signify their approval for a professional to access the specific data required or for the sending of a report (e.g. to a register). Without that approval the professional would have access only to their own records for that individual (as at present) and no data could be shared—except as required by statute. This personal control could be achieved through a password, token or biometric identifier. However if this route is rejected, the personal privacy issues and uncertainties cannot readily be resolved other than by setting up a framework of general permissions (by class of user, by specific data element, etc.) which will never be entirely satisfactory since the flexibility to respond to unexpected situations will have been removed. Even if this course is chosen, there must be a way for the data subject to 'opt out' and for their data to be withheld from sharing—although it could still be available for statistics if adequately anonymised (see above).

As far as data for administration and analysis is concerned, it should be reported in a manner that leaves identities protected. Few analysts specifically want or need the names of the individuals, but when their working database includes these data, it is difficult to ignore them and inevitably incidents will arise where confidentiality is infringed. One option would be to provide all demographic data according to analytic categories (e.g. not providing the date of birth as 19460522, but allocating the individual to the 55—65 age range). Alternatively, the demographics for an individual could all be replaced by a single identifier in reports sent for statistics, and a look up function set-up to run when an analysis is initiated to allocate the

individual to the correct group(s) as required by the specific parameters of the research.

5. Conclusion

This development has raised a number of issues. For each of them there is a trade-off between cost, effectiveness, efficiency, risk, community benefit and personal privacy—and the final decision that is reached in respect of each issue may be a function of the location in which the decision is to be applied, and the technology that is available to implement a decision in a cost-effective way.

The following issues seem to be generally significant:

1. Is it appropriate for a single public sector database of its citizens and residents to be used for more than one purpose, or indeed for all public sector purposes?
2. Where the same individuals are represented on more than one public sector database (whether achieved as in (1) above or not), should it be possible for basic data held about them to be linked—e.g. demographics—date of birth, sex, address(es), etc.?
3. Should changes in (for example) address received by one system be propagated automatically to other connected systems as in 'data validated and entered once, should then be re-used many times'?
4. Is a matrix of multi-layered permissions (e.g. by individual user, user group membership, user class, etc.) as to which personal (e.g. clinical) data elements can be accessed by which user an appropriate basis for wide area health system security? Can it be sufficiently flexible to be functional—e.g. to cope with the future as personnel change and where demands for future access cannot be foreseen?
5. Should data sharing to be implemented within the health sector to support best quality care, with continuity and integrity, but without seeking the explicit agreement of each data subject? For example, should a doctor be able to see all test results for their patient regardless of who ordered them?
6. If the above (5) is agreeable (under whatever conditions), what should happen if results that exist are not shown to a user due to specific limitations? Should they know there is a result that is not shown, or should the result simply be omitted as if it did not exist?
7. Given that there is an expectation that results will be available and linked as under (5) above, who is accountable if the linking system is temporarily unavailable at some time and an accident results from inability to share key information?
8. Should data for statistics and analysis be provided complete with identifiers attached? If not, how should the necessary cohort and longitudinal analyses be supported?