Design Principles in the Development of (Public) Health Information Infrastructures

Roderick Neame

1University of Queensland, QLD 4072, Australia

Abstract

In this article the author outlines the key issues in the development of a regional health information infrastructure suitable for public health data collections. A set of 10 basic design and development principles as used and validated in the development of the successful New Zealand National Health Information Infrastructure in 1993 are put forward as a basis for future developments. The article emphasises the importance of securing clinical input into any health data that is collected, and suggests strategies whereby this may be achieved, including creating an information economy alongside the care economy.

It is suggested that the role of government in such developments is to demonstrate leadership, to work with the sector to develop data, messaging and security standards, to establish key online indexes, to develop data warehouses and to create financial incentives for adoption of the infrastructure and the services it delivers to users. However experience suggests that government should refrain from getting involved in local care services data infrastructure, technology and management issues.

Key Words: Regional Information Management Infrastructure Design Principles

Introduction

Public Health strategies aim to improve population health and quality of life by reducing the incidence of avoidable illness, unnecessary morbidity and premature mortality: this can be achieved by analysis and identification of threats and hazards to health, as well as by early identification and containment of new syndromes and epidemics. In order to achieve these goals, it is necessary to monitor patterns of disease and of care in order to identify health priorities, to research causes of clusters of diseases and to accumulate evidence about which interventions are effective in different clinical situations. Obtaining the data to support these vital functions can be difficult, especially where there is a need for near real time data to identify health hazards (such as failing implants) and monitor the spread/patterns of epidemics – all in the context of a budget that typically demands more to be delivered using less resources.

Data on which to make such judgements may be difficult to obtain: quality and timely data even harder. Even the most basic data on what services are purchased/ provided with public funding can be difficult to obtain, so making quality, timely and cost-effective health business decisions almost impossible. Even more elusive is data on the reasons for care decisions, and the outcomes of treatments. Data about care services provided is generally
abstracted by clerks (not by those directly involved in the care encounter) and compiled into summaries and mandatory data returns, but the quality of these data often leaves much to be desired: the degree of separation between the clinical encounter and the coder reporting on it leaves room for extensive misunderstanding and misinterpretation – not to mention simple errors of abstraction and coding. Data abstracted from records by clerks, even in the best environments, is often of insufficient quality to meet the demands placed upon it\(^1\). In addition the inherent delay in reporting may be inconsistent with the needs for real time surveillance of risks: many public health reports are more than 1 year old when released, and real time data is scarce.

There is a volume published by the World Health Organisation on Improving Health Data Quality\(^2\) which contains much useful material on the topic. However it starts from a palpably false assumption, which is that everyone, including clinical staff, is dedicated to the production of high quality coded data about each and every care event and encounter. Few clinicians are even remotely interested in servicing the needs of public health information: their priorities are with the care of their patients, the enhancement of their personal diagnostic and therapeutic acumen, their research interests (if any) and their professional standing and, last but not least, their remuneration. Even so, clinical input is essential in providing the high quality data required for public health purposes and this presents a real challenge. There is plenty of analysis as to why data quality may be poor, and prominent amongst the factors is the lack of clinician involvement as well as poor working arrangements between clinical, ward, records and coding staff.

Poor quality data is reassuring, but falsely so, since it tells a story that is materially different from what exists and is happening in the field. The absence of data may be ‘better’ than poor data, simply because it does not falsely reassure, and does not divert attention from issues that are actually priorities. The US Institute of Medicine reports that many care errors and adverse incidents occur as a result of poor data and information\(^3\); but more than that, poor quality data increases costs as well as preventing measurement of performance, impeding research and analysis, and obstructing quality assurance\(^4\).

**Ways Forwards**

There are four significant considerations that provide ways of taking things forwards. Each is briefly outlined below.

**Information Economy**: Quality information suffers from being seen as an ‘add-on’ to the main activity and services for which the provider is paid. As such, it appears to lack importance and status, and this is reflected in its management at every point: the minimum possible effort is invested in reporting the data, since it is not ‘worth’ enough for anyone to pay for it. It is self-evident that quality data has a value: the logic is to separate the information about care services provided into a separate ‘economy’ which recognises the value of quality information as an entity separate from the care services themselves, and rewards those who provide it. Therefore consider splitting the payments due to those providing care services into 2 parts: one part would be for providing the service(s) to the patient; the other part would be for the provision of timely, accurate and auditable data on the reasons for and clinical data associated with the provision of the care service(s). These fees can be adjusted relative to each other in order to secure the required result. Where providers fail to furnish the required information within the allowed time period about the service(s) provided, they will receive only one part of the total potential fee.
As soon as there is a financial incentive associated with the provision of quality clinical information, one of the main obstacles is overcome: the development of an information economy where there are profits to be made will rapidly spawn new services designed to supply that need, and will assure the interest of clinicians. The financial incentives for information will drive the adoption of information management systems which can provide the required data automatically and quickly. This investment in information systems will create a marketplace where value added services provided by private enterprise will become very attractive.

Feeding back information derived from data collections is often an invaluable aid to establishing the value of data and information: such information can help drive efficiencies, promote effectiveness, reveal poor performance, identify areas of risk and generally improve competitiveness and services delivery.

**Local Valued Resources:** There are information resources that are carefully maintained at a local level because they support the needs of staff working with those patients: where information resources are valued by local staff, the information they contain will be accurate and validated. The key, therefore, is to access these data in order to generate the data required for public health purposes. For the most part the ‘key’ resource is the medical record, and it is from this document that clerks attempt to abstract data for required returns. Increasingly medical records are moving towards being held electronically in point-of-care systems, and this makes abstracting public health data automatically relatively easy. So supporting the development of electronic medical records is a priority for public health: and ensuring that any public health data developments that take place are entirely consistent with the implementation of electronic records systems is a priority. Developing a ‘public health information system’ that are incompatible with electronic medical records, whether these are introduced earlier or later, would be absurd.

**Data Aggregation:** The crucial issue is to ensure that data collected from different sources is able to be aggregated. For that to happen, there must be agreed data definitions (what each term means), agreed classifications and codings and agreed data sets to be provided in respect of each reportable situation so that the data from one encounter can be combined with the same data from many other encounters and locations. Whilst the temptation to develop ‘new’ classification and coding systems is ever present, the disadvantages of this are so extensive that the idea should be rejected out of hand.

The critical element in planning a health information infrastructure lies here. Each healthcare enterprise that implements information management technology that suits its needs can be seen as an ‘island of technology’. Those islands can be structured in many different ways using open or proprietary applications, classifications, codings, interfaces, messages and data definitions – and it has often been to the advantage of vendors to make their systems proprietary and base them on their own ‘in-house standards’. Joining up these islands is the critical challenge, and for this there has to be a clear specification of how they will be joined at technical connectivity as well as data exchange levels. The technical connectivity has been effectively by-passed since the industry has found ways of linking all types of systems because of the commercial pressure to enable linking into the Internet. The remaining issues relate to data: definitions, classifications, coding, sets, messages etc; these are where the effort has to be invested to create an infrastructure that can join the islands together, irrespective of how they function internally.
Public Health Data Warehouse: The value in any data collection will only emerge when large quantities of quality data are aggregated into a warehouse that supports sophisticated analysis. Much of the current analysis of the data is based on hypothetico-deductive research: an hypothesis is developed and the data is used to support or refute that hypothesis – which is the way research has been carried out for decades. The problem is in the nature of the hypotheses that are developed and tested: because of human cognitive limitations, the hypotheses tend to be relatively simplistic, deterministic and Boolean, of the form “IF sign A positive, AND test B positive, AND medication C negative, THEN diagnosis D”. However medicine is increasingly revealing itself to be based on relationships and associations which are more multi-factorial, fuzzy and probabilistic – none of which humans find comfortable to work or hypothesise with, although they present no problems for computers.

Developments in data warehousing and mining, driven mainly by commercial and retail interests, provide the technology for intelligent systems to analyse large data collections to identify patterns and associations that were previously unsuspected and/or unrecognised. The capacity to ‘drill down’ into the data warehouse allows these associations to be explored in greater detail. Even where there is no apparent scientific reason or explanation for a cluster of data with various common factors and data associations, the fact that it exists, and is statistically significant, is important in its own right and may suggest new avenues for study and research, and ultimately for prevention and treatment.

Health Information Infrastructure Implementations

In 1993 the first National Health Information Infrastructure went live in New Zealand (NZHIS). Some 18 years later a far bigger health information infrastructure for the UKs National Health Service, was formally abandoned in 2011. The similarities and differences between these are useful as a basis for deriving some design and development issues and principles.

1. New Zealand

The NZHIS\(^5\) was the first such national health information infrastructure: the author was chief government consultant for design, development and implementation. The system cost less than USD$5million, which was recouped in less than 1 year from retirement of legacy systems and services; there was a 2 year development period. The stated aims of the system were to support financial accountability in the context of the separation of funder and provider roles (previously funds were disbursed to providers without knowing what was being purchased), to facilitate and promote information integration between primary and secondary care, to support the national public health agenda and to allow non-government service providers and health care plans/insurers to compete for public funds and offer alternative services to the community.

A prime focus of the system was to support the public health agenda. The system was designed to gather the data required to identify community health needs, evaluate health policy, allocate resources equitably, monitor service quality and performance, and meet reporting obligations. Strong emphasis was placed on an open and contestable architecture, where a key parameter was the specification of standards for data and connectivity which were developed to act as a guide to service providers in their information systems procurements. The system created an online national healthcare user index\(^6\), a personal care summary (conditions, treatments, warnings and immunisations etc) accessible to authorised users from any location, and a minimum data set defined and to be collected for each
secondary care event. The event data was copied to the funding agency for payment management. A major emphasis was placed on data privacy, and on explaining to all parties how their data was protected: this involved legally binding agreements with users and telcos, encryption, and robust data pseudonymisation. A detailed review of the goals and focus of the system was published contemporaneously.

Since that time there have been many enhancements of that system, some initiated by government, and others by the private sector. The emphasis on data and communication standards promoted the implementation of electronic records systems and services. The non-proprietary nature of the infrastructure, together with its emphasis on standards, created a viable marketplace and encouraged many third party technology providers to offer enhanced services compatible with and leveraging off that system. Ultimately the NZHIS was dis-established in 2008 having fulfilled its developmental purpose: its functions and services were distributed amongst other government departments who took on responsibility for their operation and maintenance (eg user and practitioner indices, data warehouses, classification, terminology and data dictionary services etc).

2. UK

The UK NHS National Program for IT (‘connecting for Health’ CfH) was initiated in 2005, and was formally terminated in 2011 following a formal audit which revealed an unacceptable pattern of delays, performance problems, and extensive professional concerns as to whether the plan was deliverable. The system cost somewhere in excess of £10Billion. The aims of the system were to provide patients with more choice and control, to provide better information for patients and clinicians and thereby to deliver better care, to reduce the risks associated with care, and to provide quality information for secondary uses, especially public health.

The core planned services included delivery of electronic records (EHR) systems with detailed care records held locally and summary care records held centrally/nationally on ‘the spine’, applications for online booking of referrals (‘choose and book’ C&B) and electronic prescribing (EPS), picture communications (PACS), as well as some improvements to connectivity with greater security (virtual private network VPN) and an NHS email directory service. The spine system was intended to act as the records repository and therefore as the main resource for individual identification and those services depending upon it, as well as being the data warehouse for encounter/event reports and payments management. The plan was divided up into two parts: national services (eg the ‘spine’, the VPN and email services etc) and the local services. For the local services, a small number of local implementation service providers (LISPs) were identified, each of whom was contracted to create a system and deliver it to institutions within an allocated geographic region, so giving the end users no choice in the systems available to them – other than to decline to accept them. The CfH data privacy plan was seen as flawed from its inception and was brought into question by several experts.

In broad terms it can be seen that the core goals and the national services of this system were congruent with those of the NZ system outlined above – that is to create a central data repository, with online patient and provider indices, and online access to key personal health information, as well as a set of standards for data and communications. However the UK plan extended into additional areas, such as EHR, PACS, EPS, C&B, and what amounts to an NHS VPN: these were areas that the NZ system deliberately left blank to enable institutions to choose those services they valued (in the light of the nationally defined data and
communications standards), and to permit private enterprise to devise, develop and market such services.

**PHII Design and Development Principles**

The crucial requirement at a functional level is that the system should make possible the aggregation of data within a common data structure and format – in other words that the same terms mean the same thing to all those connected to the system, and that there is a common format for aggregation of data, including data classification and coding and the sets of data to be collected. At the same time this infrastructure enables the exchange of data ‘sideways’ between care providers and enterprises – the only difference being that there must be common standards for a wider range of data elements as well as a wider range of sets of data to be exchanged (eg tests and investigation requests and reports, administrative data on admissions, separations, transfers, pharmacy prescriptions, discharge summaries, entire electronic records exchange etc). There is no fundamental difference between the infrastructure required for data collections for public health purposes, and for data exchanges between providers: and it is vital to ensure that ‘public health’ data is not seen as different in any way, nor is it developed separately from ‘general’ health data.

Government, as the coordinator and principal source of funds, has a vital role to play in making this happen. Government must show initiative and leadership in setting standards (with the relevant professionals) in respect of the data sets to be exchanged, the message structures and formats by which they will be exchanged, the data classification and codings, and the data definitions. Almost all of this already exists in various repositories: however there are often several alternatives that could be used, and the sector as a whole needs to decide which to select for their purposes, and where there may be alternatives, options or deficiencies that need to be managed. This creates the vital piece which enables the various parts of the health sector to communicate, but it does not impose on them any requirement as to how they manage their own data internally within their ‘island’: that said it soon becomes clear that in order to make best use of the infrastructure, there are some internal data infrastructures that will align better with the external infrastructure than others.

It is here that the core information systems development principles become most relevant. These are based on the guiding principles formulated at the inception of the NZHIS project and followed throughout its implementation.

1. The system should facilitate integration of personal health records horizontally between service providers as well as aggregation vertically to ‘higher levels’ in the system, including summaries of care and preventive records as well as current personal clinical alerts and warnings (eg significant conditions and risks, important current treatments and medications)

2. The system should be based as far as possible on an open and contestable architecture and messaging infrastructure, with standards for data and communications clearly specified: proprietary systems and services should be used only where there is no practicable alternative, and even then the proprietary restrictions should be negotiated away as far as possible.

3. The communications environment should be specifically selected to facilitate and encourage third party providers to develop value-added services on top of the basic
Design Principles in the Development of (Public) Health Information Infrastructures

and national services developed by government.

4. Local information systems are the province of local management and should be selected by local management and clinicians to meet and support their needs: the ‘centre’ (health department, government etc) should be at all times aware that stepping over the threshold and becoming involved in the choice and operation of local systems greatly enhances the risk of failure where all problems can be laid at the feet of government interference, irrespective of their cause.

5. Government must take on the key role of managing the online healthcare users and providers indexes, and of promulgating (with proper consultation) data definitions, data sets and messaging standards so facilitating information aggregation and exchange; but government must refrain from developing clinical or administrative systems or imposing choices on institutions as to what systems to select and how to manage them.

6. Information privacy and systems security are not only an ethical imperative but a legal obligation, and an issue of the highest sensitivity: it must therefore be planned as an integral element of all systems and services ensuring the highest level of ethical acceptability, and these plans opened to public scrutiny. In particular the use of robust identification of staff/users and patients is essential; and robust pseudonymisation (see below) of all personalised data used for purposes other than clinical (inclusive of payment and audit).

7. All users must be enabled to connect with the system at minimal cost and with the minimum of barriers to entry, irrespective of the brand, size and platform of the internal systems they have chosen, and using the services of their own IT systems providers/support: this generally means development of a free basic API (applications program interface) that can be run on any platform, but can be fully integrated into systems as and when users decide to do so.

8. Honest and open explanations of the needs, purposes and solutions being adopted, and especially the approach to privacy and security, should be disseminated widely in formats designed for the different categories of individuals (health professionals, administrators, lay public etc). Bridges of common understanding need to be built between government and health professionals, as well as with strategic community groups.

9. Incentives for using the systems need to be incorporated. Where government funds care services, payments can be linked to provision of data, and speed of payment can be linked to the speed with which data is provided. The unique national patient identifier can be required to substantiate all basic claims for payment; the prescribed minimal data set relevant to the clinical situation can be required to support claims lodged electronically for the full payment; and electronic reimbursement can be made the same day as claims are received and validated.

10. Updates to data definitions, sets, classifications and coding systems, message definitions etc must be negotiated with the sector and published some considerable time ahead of their mandatory introduction, so that institutions, their IT services and systems developers have sufficient time to incorporate these into local systems.
As a brief observation, it would seem that the UK NHSs CfH project definitely adhered only to principle 1 above: it seems likely that it breached principles 8 and 10 above, and it is clear that it breached principles 2 – 7. Principle 9 is probably irrelevant in the context of the operational management of the NHS.

**PHII Benefits and Risks**

The benefits from a PHII development are twofold. One benefit is that information can be exchanged between providers caring for the same patient, improving continuity and integrity of care, and allowing patients to choose where they go for care services, rather than being ‘locked-in’ to an institution which holds their medical records. The other main benefit is the aggregation of data into warehouses that permits all types of cross-sectional and longitudinal studies to be undertaken to analyse incidence of diseases/syndromes, immunisation and prevention status, best care practices, previously unknown associations between entities, etc. All of this will become invaluable as the progressive move is made into greater use of artificially intelligent decision support and alerting systems, which rely heavily on a comprehensive and up-to-date knowledgebase derived from the evidence that is abstracted from the data warehouses.

There is a potential risk to patient information privacy. All data passing across public networks can be protected from eavesdropping by strong encryption, using a technology appropriate to the risk, but migrating progressively towards a secure public key infrastructure (PKI) encryption environment.

For the most part it is quite unnecessary for the identity of the patient to be attached to data used for research purposes: the personal identifying elements can be replaced with a cipher, a process sometimes known as ‘pseudonymisation’. This is effective only where it is robust, and there is no ready access to enable users to re-establish the identity of the individual – although as in the NZHIS a ‘key in escrow’ arrangement can be made so that in the event of, for example, a serious problem being identified that could threaten the well-being of individuals (eg a faulty implant), a decision can be made at top level to apply the key solely to re-identify those affected and advise their care provider(s) of the potential risks.

**Information Feedback**

The value of health information and evidence lies in making use of it to improve community health status, to inform and educate both clinicians and patients, and to get the best possible value for every health dollar that is spent. Generating data is all well and good: but using it effectively is vital. The research shows that those providing data do so more willingly and conscientiously if they get something back from their efforts, so feeding back useful information to the workforce is all important. Tables of statistics for many people have little impact: graphic representations of the data (pie-charts, histograms etc) often mean much more to the recipients of the information, and it is only if they understand the data that they will look to modify their behaviour appropriately.

Timely data is the most useful, so providing updates on current outbreaks of disease and on newly identified syndromes is vital. Most competitive services welcome comparative feedback identifying strengths, weaknesses and opportunities for improvement. The use of charts which place the performance of each service provider/unit in the context of the performance on the same parameters of all similar service units (all being anonymised), gives
a clear idea of where there is cause for concern as well as for self-congratulation. Crucially as cost-effectiveness becomes the new driving force in health service delivery, it will be vital to compare unit performance based on their adherence to best practice guidelines and on overall costs for each clinical entity.

Timely feedback is essential. Where feedback is delayed, bad results can and usually are dismissed as out-of-date and ‘changes have already been made’ to improve performance. The goal must be real time feedback preferably whilst the patient is still in care identifying those individuals where care costs are out of control, and clinical parameters/outcomes are sub-optimal so that lessons can be learned before it is too late.

But it is just as important to engage the community in this feedback process, alerting them individually and as a community to risks and hazards, to better and worse performing care service units, to epidemics, to the need for appropriate preventive care and much more. Patients have to make informed decisions about their own health and the way in which they can make best use of the available services: they can only do this if they are well informed about risks and options.

**Where Technology is Less Advanced**

The impact of this sort of approach on care service providers depends on the level of technology they have access to. Those with no automation, not even an office computer, will be able to provide their data for an initial period on paper forms – but this should be phased out with incentives to move to a higher level of technology. Where there is basic office automation – just a computer connected to the internet – providers will be able to use the free API to submit the required data in support of their claims for payment. Those with more advanced systems will be able to use the infrastructure specifications to have their IT staff develop an interface between their systems and the API to enable fully automatic submission of data and claims.

Systems developers and providers will have a clear information infrastructure definition to guide their development of next generation systems. It is vital that the full set of required data elements for each clinical situation are collected within the software and coded using the agreed classification and coding system in order for the link between the systems and API to be easy to engineer.

Once the infrastructure has been clearly specified, and there is a clear marketplace, it does not take long for entrepreneurs to identify a range of value-added commercially viable services that can be developed for health sector users, compatible with the infrastructure and offering further performance enhancements and benefits to users, so effectively further embedding the use of these systems in the sector. In this way the relatively small investment of the government in infrastructure leads to a much larger investment by the private sector in an expansion of the environment.

**Conclusion**

Public health information management must be developed as part of a general health information management strategic plan: they need to be developed side-by-side to ensure complete consistency and compatibility. Strategies need to be implemented that can engage the interests of clinicians in the provision of quality, timely information: associating the provision of information with financial incentives is suggested.
Privacy concerns always emerge as a key issue in such information infrastructures and the data repositories associated with them. Both longitudinal and cross-sectional research studies can be conducted on pseudonymised data without any breach of personal privacy, although maintaining a decryption key-in-escrow may be a wise precaution.

Data warehouses and their tools for data mining will bring considerable added value to the data collections, and analysis using neural networks will quickly identify patterns and association in the data that human analysis cannot readily discern. These data collections will be invaluable in determining best quality practices and providing the knowledgebase for artificially intelligent systems in healthcare. Feeding back information abstracted from such analysis to those providing the information, as well as to the public, will be important in ensuring the continuing cooperation of clinicians and patients alike, and in ensuring practitioner adherence to best quality care protocols.

Distilling down the 10 principles outlined above, the big issues, based on a wealth of practical experience, appear to be:

- That the issue of personal information privacy protection, both relating to patients and to care providers, is addressed thoroughly and planned for meticulously in the context of both the law and highest ethical principles, and laid open to public scrutiny
- That government takes a leadership role and defines the required standards for data interchange (data and messages), as well as creating the requisite ‘back-end’ services to support the system (eg online identifiers, data collections/warehouses etc) – all based on open and non-proprietary standards, and with minimal barriers to adoption and use
- That government does not impose systems or services on clinical service providers and enterprises, thereby infringing their autonomy, but having defined the infrastructure and created incentives for its use, then leaves commercial vendors to develop and market value-added services that leverage off that infrastructure.

**Limitations**

The 10 principles outlined above have been derived empirically: there may be others that are equally relevant, but have not yet been identified; and the 10 that have been outlined will likely benefit from further refinement and modification. Because of the size, complexity and expense of such major projects, however, it is difficult to envisage that there will be many experiments conducted specifically to test the principles. However it may be that where such infrastructures are being planned and developed, those involved may reflect on the principles, decide in advance which to adopt and which to dismiss, and subsequently review their progress, and difficulties, in the light of these principles.

Some of these same principles might be applicable to the many smaller (eg enterprise wide) systems integration projects that arise as enterprises acquire new facilities and seek to integrate them into their existing care and billing infrastructure. However for the most part these projects tend to revolve around pragmatic decisions as to how to extend existing systems (good or bad) to embrace new members, rather than exploring how best to link together multiple islands of technology each of which has as much merit as the next, and at the same time to develop the resources required for the ‘public good’ that support better management of public health.
Conflicts of Interest: None

Correspondence

Roderick Neame, BA,MA,PhD,MB,BChir,FACHI
Health Information Consulting Ltd,
16 Glen Eden Court, Flaxton, QLD 4560, Australia
Email: roddyneame@hic-ltd.com

References