

White paper
on
Standardization of Health Data
(Version 0)

Vicky Fung

Manager (Health Informatics)

Hong Kong Hospital Authority

September 2006

Acknowledgements

The author would like to express a heartfelt thanks to Dr N T Cheung and Dr C P Wong for their guidance and advices in the preparation of the report. This report cannot be produced without the shared experience of all clinicians who have been involving in the development of the Clinical Management System, and various health informatics and information technology colleagues who have been working in the health information systems at the Hospital Authority. The author is in debt to all Health Information & Records Managers for their advices in health information management. Special thanks is given to the following colleagues who advised on the report content : Ms S C Chiang, Mr Frank Chung, Mr Warren Cheung, Mr James Tan and Mr William Chan. The author appreciates the understanding and support given by the colleagues of the Health Informatics team when the report is written. Last but not the least, the author would like to thank all members of the Data Sharing & Standards Working Group for their support to the project : Dr C P Wong, Dr Heston Kwong, Ms Conser Lee, Mr Owen Wong, Mr K C Chan, Dr K T Chan, Dr C P Ho, Ms Manbo Man.

Table of Content

1.	Standardization of Health data	4
2.	Computerization of Healthcare Sector in Hong Kong.....	5
3.	Identifying a Person	7
3.1.	Information Source	7
3.2.	Registration Practice.....	7
3.3.	Data Entry Accuracy	8
3.4.	Information Relating to A Person	8
4.	Identifying Healthcare Providers / Specialty / Institute	10
5.	Record Structure	11
6.	Terminology Standards.....	13
6.1.	Introduction to Terminology Standards.....	13
6.2.	Drug Data	14
6.3.	Laboratory Test Codes	15
6.4.	Imaging Data	15
6.5.	Other Clinical Data.....	16
6.6.	Issues with Terminology Standards	18
7.	Messaging Standards	21
8.	Knowledge Representation	23
9.	Data Privacy, Confidentiality & Security.....	25
9.1.	Introduction	25
9.2.	Developing Standard Practice Guides	25
9.3.	Applying Technology Standards	26
9.3.1.	Multi-factor Authentication.....	26
9.3.2.	e-Certificate	26
9.3.3.	Digital Signature	27
9.3.4.	Cryptography	27
9.4.	Standards Body on Data Security	27
9.5.	Education on Data Privacy & Security	28
10.	Management of Data Standards.....	30
11.	Conclusion.....	32
	Glossary	33
	References	35
Appendix I	Adoption of Standard Terminologies in Hong Kong Hospitals.....	36
Appendix II	ASTM E31.20 Standards on Privacy and Security	37
Appendix III	CEN Working Group 3 on Security, Safety and Quality.....	38

1. Standardization of Health data

- 1.1. Data standardization is the basis for accurate and efficient communication. Healthcare service is contributed by multiple carers at different times at various locations. Data standards ensure information is interpreted by all users with the same understanding. It facilitates comparability of data and interoperability of systems for data captured at applications built on different platforms. It supports reuse of data, improves the efficiency of healthcare services and avoids errors by reducing duplicated efforts in data entry. Data standardization assists in protecting and promoting population health through early detection of abnormal patterns.
- 1.2. In healthcare environment, data standardization refers to i) the data content; ii) the terminologies that are used to represent the data; iii) how data are transmitted across various platforms/applications; and iv) how knowledge, e.g. clinical guidelines, protocols, decision support rules, are represented in the health information system ⁽¹⁾.
- 1.3. The basic elements for data sharing requires standardization in the areas of identification, record structure, terminology, messaging and knowledge representation. Issues of data privacy and management of standards are also important.
- 1.4. The following will review the current status of computerization and standardization of health data in Hong Kong. The report will further discuss the basic elements and propose steps to be taken to facilitate data sharing amongst healthcare providers.

2. Computerization of Healthcare Sector in Hong Kong

- 2.1. Computerization of healthcare records in Hong Kong mainly concentrated in public hospitals and their clinics. The Hospital Authority (HA), managing all public hospitals, started to implement the patient administration systems in 1991. All patients attending HA institutes are registered in the Hong Kong Patient Master Index (HKPMI). In 1995, the Clinical Management System (CMS) was first implemented in the public hospitals. At present, important clinical data are being entered into CMS workstations by nearly 4,800 doctors, 20,000 nursing staff and 4,800 allied health professionals. These data include allergy history, diagnoses, procedures, discharge summaries, operation records, special investigation reports, laboratory results, radiology reports and images, medications ordered / dispensed, and other specialty based clinical data. Starting from 2000, HA healthcare professionals can view all captured electronic data created at any HA institute via the electronic Patient Record System (ePR) at any HA workstation. A pilot project was launched in April 2006 to share the ePR with registered private practitioners with the patient's agreement.
- 2.2. Leveraging the CMS, the HA built the e-SARS Reporting System to monitoring the SARS outbreak in 2003. e-SARS, together with the address table developed by the Police Department and the contact tracing system developed by the Department of Health (DH), played a key role in managing the SARS outbreak. Starting from 2006, HA doctors are reporting notifiable and communicable diseases to DH electronically via the Notifiable Disease and Outbreak Reporting System (NDORS).
- 2.3. In March 2005, DH launched the Central Notification Office On-line to provide an electronic channel for private doctors to report notifiable and other infectious diseases of public health concern. In addition, there are several projects managed by the DH to facilitate monitoring the public health, e.g. the Public Health Information System, Cervical Screening Information System. All these initiatives aim to provide a convenient and secure electronic platform to enhance surveillance of diseases such that actions can be taken as soon as possible to protect the health of the Hong Kong community.
- 2.4. A recent survey done by the Hong Kong Medical Association ⁽²⁾ indicated that there is an increasing trend / interests in using computers at the private health care sectors. Over 85% of them used computers at work, mainly writing letters, internet access and indexing patient demographic data. Around 49% of them also used computer for drug labeling, and 43% of them were documenting patient clinical data in computer.

- 2.5. The Working Group also surveyed various hospitals on the adoption of data standards. However, only HA and one private hospital responded. Please refer to Appendix I. Given the low response from the private sector, one could hardly draw a conclusion on how well health data standards are being adopted in the Hong Kong healthcare system. However, the survey indicated that :
- 2.5.1. The HA hospitals have adopted a number of standards in her clinical information systems. For some, in particular, terminology, standards are only served as a reference to the captured data instead of directly applying the standard as the captured data. For example, the HA Clinical Vocabulary Table (HACVT) is modified from the ICD 9 CM to support capturing diagnoses and procedures. One-third of the HACVT content is locally added terms. The Drug Table being used by the HA is also built from the British National Formulary (BNF) and added drug names that not included in the BNF.
- 2.5.2. Some standards are being used at HA hospitals and also some private hospitals, e.g. using Hong Kong Identity Card / Birth Certificate number to register their patients; referencing diagnoses and procedures to ICD 9 CM; using DICOM to represent imaging data.
- 2.5.3. The HA is also adopting standards that become more popular overseas or those developed in recent years, e.g. SNOMED CT, ICF, LOINC.

3. Identifying a Person

All persons must be uniquely and accurately identified for establishing a womb-to-tomb personal health record. That includes 3 elements : accurate information source, standard registration practice, and accurate data capturing.

3.1. Information Source

3.1.1. All Hong Kong citizens are uniquely identified. Hong Kong Identity Card (HKID) number is issued to children over 11 years old by the Immigration Department. It is a reliable and convenient means to uniquely identify Hong Kong citizens. Identification of children under 11 years old relies on their Hong Kong Birth Certificate number which will eventually become the person's HKID number. In the survey, both HA hospitals and the other private hospital are using HKID and Birth Certificate number to register their patients.

3.1.2. Linking a person's birth data with one's subsequent health record is relied on the parents' initiative to provide the baby's Birth Certificate number to the birth hospital. The baby's surname is not reflected on the Birth Certificate. Yet, there are various reasons that the father's surname would not be used as the baby's one, e.g. single mother, ritual of some ethnic groups. However, the Immigration Department indicated that amendment to the Birth Certificate and provision of Birth Certificate number to the birth hospital would require a revisit of the existing law. Not all persons will present one's HKID card on registration. Fake HKID cards are on the news on and off. Once a HKID number is shared by more than one person, the health data attached under that number will be mixed up as well. It is difficult and time consuming to identify and verify which piece of clinical data belongs to which person.

3.1.3. Considering the resources required to develop and maintain a unique personal identification system, it is worthless to replace the HKID number by another unique identifier. Yet, there is a need to explore means to strengthen the identification and authentication (please refer to Chapter 9) of a person using new technology.

3.2. Registration Practice

3.2.1. The Chinese culture, e.g. husband's surname not shown on the HKID card, date of birth in lunar calendar, may confuse the registration staff. Other concerns include

persons with trans-sexual operation, newborns of multiple pregnancy... Without standard practice to register a person, this would result in repeatedly updating of a person's identifier.

- 3.2.2. The increase in number of visitors from mainland China signifies the need of reviewing the existing system in just using traditional Chinese characters to identify persons whose identity document only bear one's name in simplified Chinese characters.

3.3. Data Entry Accuracy

- 3.3.1. Typing error of a person's identifier is one of the major reasons that the system cannot link one's health record. The face data that stored in the smart HKID card provides a reliable source of information for capturing one's identifier if that is made available to the healthcare industry.

3.4. Information Relating to A Person

- 3.4.1. In addition to personal identifiers, other data that would assist in monitoring public health, e.g. address, ethnicity, also need to be standardized. Special attention should be given to those vulnerable groups, e.g. elderly home, people living in institute. Currently, the HA is using the elderly home code table provided by the Social Welfare Department to identify patients sent from elderly homes.
- 3.4.2. The address table shared by the Police Department played an important role in combating SARS. The Office of Government Chief Information Officer (OGCIO) is currently working with the concerned parties in developing a standard address table to capturing of quality address data as well as to facilitate spatial analysis of the captured address data. This will serve as a sound foundation for all sectors, including health care sector in both patient and service management.

3.5. Recommendations

- 3.5.1. *To adopt Hong Kong Identity Card number / Birth Certificate number as the unique identification number for healthcare seeker.*
- 3.5.2. *To register and verify a person using biometric data, e.g. reference to the biometric database that kept by the Immigration Department.*

- 3.5.3. *To liaise with the Immigration Department to support the healthcare industry accessing the face data that is being stored in the smart HKID card.*
- 3.5.4. *All babies should be properly identified with their surname and given name in the birth certificate.*
- 3.5.5. *To discuss with the Immigration Department to share the Birth Certificate number with the birth hospital to facilitate building a lifelong personal health record for Hong Kong residents.*
- 3.5.6. *To educate the public on one's responsibility to provide accurate personal identifier during healthcare attendance for linking previous health records.*
- 3.5.7. *To study how the OGCIO standard address table would be implemented in health care sector.*
- 3.5.8. *To develop practice guidelines on registration of a patient and how clinical data should be verified when personal identifier is mixed up.*

4. Identifying Healthcare Providers / Specialty / Institute

- 4.1. Identifying healthcare providers is essential for future communication amongst healthcare providers on a person's health data. These include all healthcare providers who could involve in the care of a person, e.g. doctors, nurses, allied health professionals, professionals working in laboratories, radiology unit, pharmacies, persons working in the elderly home, or even the person himself, his family. All essential administrative and organizational information relating to healthcare providers, e.g. their affiliated institute, their specialty, should also be standardized. This assists in tracing the related specialist in case the healthcare provider in-charge cannot be contacted at a particular point of time.
- 4.2. All healthcare professionals need to register themselves to the relevant professional body so that they are allowed to practice in Hong Kong. This provides a reliable source to identify the profession of the healthcare provider.
- 4.3. Another important issue is to authenticate the healthcare provider (please refer to Chapter 9). Improper practice of using computer may affect how a healthcare provider is being identified, e.g. entering data into system that previously not logged out. This affects subsequent communication of a person's health data, and also may affect a person's data privacy.
- 4.4. *Recommendations*
 - 4.4.1. *To liaise with various professional bodies to establish a central database to uniquely identify the healthcare professionals and healthcare institutes.*
 - 4.4.2. *To adopt multi-factor authentication to verify the identify of the healthcare provider.*
 - 4.4.3. *To explore new technologies, e.g. biometric mouse, digital signature, to uniquely identify and authenticate a healthcare provider.*

5. Record Structure

- 5.1. Defining the record structure facilitates the organization of clinical information and its subsequent retrieval. The amount of health data being captured varies due to a number of factors, e.g. different penetration of information technology at various sectors, different approaches in implementing information system for healthcare sectors that have already introduced computerized health record. A clear direction is required so that the health sector can focus their development of electronic health record, including the essential record content, and the data standards, e.g. data definition, data format.
- 5.2. In overseas, there are standards developed on the record content. The Continuity of Care Record (CCR)⁽³⁾ (see Table 1) is developed and maintained by the ASTM E31 Committee. It recommends the core data set to be sent to the healthcare provider when there is a referral / transferred.

Table 1	Content of Continuity of Care Record
1.	Header, or Document Identifying Information <ul style="list-style-type: none">✓ referring or "from" clinician✓ referral or "to" provider✓ document date✓ purpose for creating the document✓ reason for referral
2.	Patient Identifying Information
3.	Patient's Insurance and Financial Information
4.	Health Status of the Patient <ul style="list-style-type: none">✓ Diagnoses, Problems✓ Adverse Reactions/Alerts✓ Current Medications✓ Immunizations✓ Vital Signs✓ Laboratory Results✓ Procedures/Assessments✓ Health Status
5.	Details on the patient-clinician encounter history
6.	Care Plan

5.3. ASTM also recommends standards on healthcare documents and the various sections in a person's health record in its E1384-02a Practice for Content and Structure of the Electronic Health Record (EHR).

5.4. *Recommendations*

5.4.1. *To define a minimum dataset for various types of records, e.g. inpatient, outpatient, accident & emergency visits, and for transferring / referral / discharge.*

5.4.2. *To develop the data format for the recommended minimum dataset.*

6. Terminology Standards

6.1. Introduction to Terminology Standards

- 6.1.1. Terminologies are medical terms and concepts used to describe, classify, and code the data elements and data expression languages and syntax that describe the relationships among terms/concepts ⁽¹⁾. Terminology is used to record clinical information; to facilitate the storage of clinical information; to support sharing and reuse of clinical information; to support efficient query formulation; to create a natural language output from manual structured input; and to support the application of decision support algorithms ⁽⁴⁾. Cimino listed the desiderata of a medical terminology ⁽⁵⁾ (see Table 2):

Table 2 Desidarata of Medical Terminology	
i.	complete coverage of domain-specific content
ii.	each concept in the vocabulary should have a single, coherent meaning
iii.	meaning of concepts do not change with time, view, or use, and concept cannot be deleted from the terminology
iv.	each concept is uniquely identified with meaningless identifier
v.	a concept should be put under more than one hierarchy, where appropriate
vi.	concepts have explicit formal definitions
vii.	residual categories in traditional classifications, e.g. not elsewhere classified, should not be included in a terminology as it will change with time
viii.	terminology supports concepts at multiple granularities, as required
ix.	there is a consistent view of representation of concepts at various levels / hierarchies
x.	concepts must evolve with change in knowledge
xi.	concepts should be able to represent the context of the record
xii.	there is mechanism to recognize concepts of same meaning as same concept could be expressed in different ways

6.1.2. There are various types of terminologies developed for different domains of health data. The following introduce the most commonly used ones.

6.2. Drug Data

6.2.1. GS1 Hong Kong is developing Central Drug Repository (CDR) to adopt a standard naming convention for drugs being used in HK across various parties, e.g. manufacturers, distributors, drug retailers. The system development was completed by May 2006. 2 major commercial pharmacies planned to implement the CDR. Penetration of information technology at small drug stores is still limited. The development of CDR highlights the importance of standardizing drug data to facilitate tracking dispensed drugs. However, as a drug can be manufactured by different companies, it is important to reference to the brand name to the drug's generic name, and to identify the ingredients to facilitate development of knowledge based information such as drug interactions, allergies, and contraindications to ensure patient safety.

6.2.2. Today there are a variety of drug standards, each developed for a different purpose. Unfortunately there does not appear to be a single standard which encompasses all requirements. For instance, the Hospital Authority's Drug Table is built to support the ordering and dispensing of medications; MedDRA is a terminology designed to support the classification, retrieval, presentation, and communication of medical information throughout the medical product regulatory cycle; and the First Data Bank proprietary drug classification is designed primarily to support clinical decision support.

6.2.3. Other challenges facing drug terminology is that same drug name of a particular manufacturer could be of different preparations when it is sold at different places. Individual doctor could prepare his own formula, e.g. ointment, syrup, which is not available anywhere. This formula may not be known to others but the doctor himself only.

6.2.4. British National Formulary (BNF) ⁽⁶⁾ is published jointly by the British Medical Association and the Royal Pharmaceutical Society of Great Britain. The BNF provides guidance on prescribing, dispensing and administering medicines with special reference to their uses, cautions, contra-indications, side-effects, dosage and relative costs. Drugs are grouped according to the body systems that it acts on. Thus, if a drug serves several purposes, then it would be included under all body

systems that it acts on under different identifiers. The HA and the surveyed hospital are using BNF. The HA has built another unique identifier for the drug and mapped it to multiple BNF identifiers, if applicable.

6.2.5. RxNorm ⁽⁷⁾, a standardized nomenclature for clinical drugs, is developed by the National Library of Medicine (NLM) of US. It links the branded & generic name of a drug to its active ingredient, strength (e.g., 120 milligrams) and dosage form (e.g., tablet) of a particular drug. RxNorm's standard names for clinical drugs are connected to various drugs terminologies used in commercially available drug information sources, e.g. First Databank, MediSpan. These connections facilitate the interoperability among the computerized systems that record or process data dealing with clinical drugs. It was selected as one of the core terminologies to be used in US ⁽¹⁾.

6.2.6. MedDRA ⁽⁸⁾ is developed as an initiative of the International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH). It is maintained by the MSSO (Maintenance and Support Services Organization), an organization that reports to the IFPMA (International Federation of Pharmaceutical Manufacturers and Associations; www.meddramsso.com). MedDRA is a medical terminology designed to support the classification, retrieval, presentation, and communication of medical information throughout the medical product regulatory cycle at both the pre- and post-marketing stages. This terminology includes terms for safety-related events and adverse drug reactions.

6.3. Laboratory Test Codes

6.3.1. Logical Observation Identifiers, Names and Codes (LOINC) ⁽⁹⁾ was initiated in 1994 by the Regenstrief Institute in US for coding laboratory tests. LOINC is a naming standard for 'data fields'. Its attributes describe how data is collected, and specific nature of the involved observation, e.g. type of property (e.g., mass concentration), timing (e.g., 24-hour specimen), specimen (e.g., urine), data type of the data field (e.g. narrative, ordinal). To date, LOINC also includes clinical observations, e.g. head circumference, blood pressure. It was selected as the standards for reporting laboratory results amongst all federal health care services agencies in US ⁽¹⁾. HA is also using LOINC to identify her laboratory test codes.

6.4. Imaging Data

6.4.1. The Digital Imaging and Communications in Medicine (DICOM) ⁽¹⁰⁾ is maintained by the National Electrical Manufacturers Association (NEMA). It facilitates the distribution and viewing of medical images, such as computerized tomographic scans, magnetic resonance imaging, and ultrasound. DICOM defines the essential components to capturing / storing / communicating images, e.g. data structure, semantics, message exchange protocol, network support required for exchange message, storage requirement, display function, security. To reduce image size, the DICOM image data can be compressed (encapsulated) as lossy or lossless format. It was selected as the messaging standards amongst all federal health care services agencies in US ⁽¹⁾. The survey done by the working group indicated that both HA and the returned hospital are using DICOM for their imaged data.

6.5. Other Clinical Data

6.5.1. There are various types of terminologies covering various domains of clinical concepts. However, no single terminology can cover all aspects of clinical concepts. The following introduce a few terminologies that are more popular at local / overseas health sector.

6.5.2. International Classification of Diseases (ICD) ⁽¹¹⁾ was first adopted in 1900 for collecting statistics of deceased data. It is maintained by the World Health Organization (WHO) for collecting both morbidity and mortality statistics across countries. It mainly covers diagnosis, symptoms, and external causes of injury. The latest version is the tenth revision (ICD 10). Some countries modified the ICD to facilitate billing patient's hospital stay, e.g. ICD 9 CM at the US, ICD 10 AM at Australia. The ICD 9 CM and ICD 10 AM also include a list of procedures that is developed by their own country. All Hong Kong hospitals report their mortality data and inpatient morbidity data in ICD 10 to the Department of Health since 2001. Given the main purpose of ICD is for collecting morbidity and mortality statistics, it does not have the breadth and depth to cover all contents of a clinical record ⁽¹²⁾.

6.5.3. The International Classification of Primary Care (ICPC) ⁽¹³⁾ was first published in 1987 by the World Organization of National Colleges and Academic Associations of General Practice/Family Doctors (WONCA). It facilitates the capturing of reasons for encounter, diagnoses or problems, and process of care at primary care setting. However, the terminology is not appropriate at the secondary and tertiary care setting. The HA is now using ICPC version 2 at her general outpatient patients, and the surveyed hospital is using version 1.

6.5.4. International Classification of Functioning, Disability and Health (ICF) ⁽¹⁴⁾, was developed in 2001 and is maintained by the WHO. It classifies a person's functioning and disability due to the consequences of disease based on one's body functions and structures, activities and participation and interaction with environmental factors. ICF aims to understand and measure health outcomes. Its adoption is still limited. For practical use, the HA maps ICF to their own local terms instead of directly applying the ICF terms as an interface terminology.

6.5.5. Systematized Nomenclature of Medicine, Clinical Terms (SNOMED-CT) ⁽¹⁵⁾ is maintained by the College of American Pathologist (CAP). Consisting of over 360,000 concepts, SNOMED CT is currently the largest single clinical terminology. Apart from diagnoses, procedures, drugs, specimens, it also includes other areas that could be of importance in building an electronic health record. SNOMED is continuously evolving to include other terminologies into it, e.g. nursing terminologies that used in US. It also receives subscribers' comments and may incorporate their requests in the SNOMED CT releases after review. In 2006, 6 countries (Australia, Canada, Denmark, Lithuania, the United Kingdom and the United States) joined the SNOMED CT Standards Development Organization. Negotiation with the WHO is underway on whether to adopt SNOMED CT as a standard for all countries in representing clinical data. It was selected as one of the core terminologies to be used in US ⁽¹⁾. HA is using SNOMED version 3 for her histology data, and is planning to migrate her clinical data to SNOMED CT. The major differences between SNOMED and the other types of clinical terminologies include :

6.5.5.1. covering other areas of clinical data instead of one single domain as in the other clinical terminology

6.5.5.2. defining clinical concepts through the relationship between concepts to clarify the meaning of it

6.5.5.3. adopting a polyhierarchical structure to facilitate data retrieval

6.5.5.4. supporting capturing of clinical concepts up to the granularity as desired

6.5.5.5. supporting compositional terminology – building new concepts from the existing ones

6.5.5.6. using a meaningless identifier to facilitate expansion

6.5.5.7. allowing unlimited expansion of any hierarchy under SNOMED CT

- 6.5.6. Current Procedural Terminology (CPT) ⁽¹⁶⁾ is developed and maintained by the American Medical Association (AMA) for the purpose of insurance claim on procedures performed at the outpatient setting. In Hong Kong, some insurance companies are using CPT as a reference for insurance claims.
- 6.5.7. Medical Subject Headings (MeSH) ⁽¹⁷⁾ is a controlled vocabulary thesaurus. It was developed and is maintained by the National Library of Medicine (NLM) of US. MeSH terms are arranged in both alphabetic and hierarchical structure. Its applications include indexing articles from biomedical journals, cataloging books, documents, and audiovisual materials acquired by the Library. The HA is using MeSH to index her library collections.

6.6. Issues with Terminology Standards

- 6.6.1. There are various standards bodies around. Some areas, e.g. clinical diagnosis, are better developed. Some are not widely adopted. Standards of the same domain compete with each other. Some of them complement each other to facilitate sharing health data amongst different organizations. No single terminology can cover all domains that are required in the healthcare field. Consider the complexity of health data, it would not be cost-effective to develop another separate terminology standard for the Hong Kong healthcare system.
- 6.6.2. Overseas and local experience indicates the need of local standards for local requirements. The HA is monthly updating its Clinical Vocabulary Table (HACVT) which is used for capturing diagnoses/procedures in CMS. As discussed earlier, both US and Australia also develop their own system in capturing diagnoses and procedures. There should be a similar mechanism to develop and maintain a set of standard terminologies that are applied in Hong Kong healthcare community.
- 6.6.3. In an ideal world, standard terminologies should be adopted and applied in the health information system directly. Yet, this situation is often restricted by the design of individual systems that arise from preferences of local healthcare providers, and the workflow at individual settings. Mapping of local terms to standard terminology is unavoidable. Mechanism should be set up to ensure the accuracy of the mapped terms.
- 6.6.4. Healthcare institutes are required to report data using different terminologies for various purposes at the request of individual authorities, e.g. charging, reporting

mortality data. Mapping of health record data which is often more detailed, to a terminology of less granular level is unavoidable. This implies irrevocable data loss. The complicated data capturing rules in some terminologies, e.g. ICD, do not allow a simple mapping in some situations. This increases the challenges that facing for executing logics in the computer systems based on a particular terminology.

6.6.5. Health data standards evolve with the advancement of health science. SNOMED releases its update every half yearly. ICD 9 CM, ICD 10 AM are updated on yearly basis. Updates could include adding additional concepts and / or obsolete redundant concepts or error concepts. All these updates could impact the already captured data and would affect the continuity of data for ongoing care or trend analysis.

6.6.6. In overseas, there are authorities to coordinate the above activities, and to study the impact of changes in terminology versions. For example, the National Centre for Classification in Health (NCCH) ⁽¹⁸⁾ of Australia publishes the ICD 10 AM, conducts studies to identify the appropriate terminologies for various sectors of the Australian healthcare system, establishes relevant guidelines relating to the use of terminology, and conducts trainings on issues relating to classification of health. The NLM of US conducts and supports research in natural language processing to extract usable and meaningful information from biomedical text. The Institute of Medicine ⁽¹⁾ also recommended the NLM to be the responsible entity for distributing all national clinical terminologies that relate to patient safety and for ensuring the quality of terminology mappings in US.

6.7. *Recommendations*

6.7.1. *To select a core set of terminology standards that meet Cimino's desiderata for drug, imaging, diagnosis, procedures, laboratory test as a basis for local healthcare sector.*

6.7.2. *To develop mechanisms to :*

6.7.2.1. *support extension of standard terminology to meet requirements of local healthcare sector*

6.7.2.2. *ensure quality of mapped data to be provided for various purposes, e.g. reporting mortality & morbidity, charging*

- 6.7.2.3. *ensure there is a continuity of dataflow to support healthcare and trend analysis from updates of various standard terminologies*

- 6.7.3. *To continuously review the development of various international standards that could have impact on sharing of health data*

7. Messaging Standards

- 7.1. Health Level 7 (HL7) ⁽¹⁹⁾ is a healthcare application protocol developed in 1987. It is now accredited as a standard by the American National Standards Institute (ANSI). HL7 is a widely adopted messaging standard in the health sector for the exchange, management and integration of electronic healthcare information. "Level Seven" refers to the highest level of the International Organization for Standardization (ISO) communications model. The seventh level supports such functions as security checks, participant identification, availability checks, exchange mechanism negotiations and, most importantly, data exchange structuring. The latest approved version is HL7 version 2.5 which was selected as the messaging standards amongst all federal health care services agencies in US ⁽¹⁾. In a recent project of developing electronic birth return, the HA and the other private hospitals are using HL7 version 2.3 as the messaging standard for transmitting data to the Immigration Department. HA is also using HL7 for communicating health data with various external organizations, e.g. Department of Health, and servers of vendor machines.
- 7.2. HL7 is further developing standards in data exchange and has developed Reference Information Model (RIM) ⁽²⁰⁾ which is the basis for HL7 version 3. RIM provides an explicit representation of the semantic and lexical connections that exist between the information carried in the fields of HL7 messages. It is developed with the aim to provide a consistent view of the data that HL7 moves, and to address the data's relationship to other data. RIM defines healthcare interactions (clinical, administrative and financial activities) using an object-oriented approach. However, HL7 2.x is not directly convertible to HL7 v.3. All existing exchange messages have to be revamped if HL7 v.3 is adopted. Its unnecessarily verbose XML style also hinders its adoption in the field ⁽²¹⁾.
- 7.3. Other HL7 initiatives include Clinical Document Architecture (CDA) which was approved as ANSI standard in November 2000. CDA provides an exchange model for clinical documents, e.g. text, sound, and multimedia, by using extensible markup language (XML).
- 7.4. Apart from data entered by healthcare providers, there are also data captured by medical devices at the bedside. These are mainly physiological parameter measurements and device settings captured by devices used at intensive care unit, operating room, and emergency room. These data could be captured at snapshot, or they could be continuous feeding from the medical device.

7.5. IEEE (Institute of Electrical and Electronics Engineers) 1073 ⁽²²⁾ provides communication between the bedside devices in Intensive Care Unit, Operating Room and Emergency Room with the server. The trend of using wireless equipment is increasing and there is a need to consider a messaging standard for data generating from the medical device if these data are to be incorporated into a person's health record. It was selected as the standards for medical devices amongst all federal health care services agencies in US ⁽¹⁾.

7.6. *Recommendations*

7.6.1. *To adopt HL7 v. 2.5 as a messaging standard for sharing health data*

7.6.2. *To further review the requirement of adopting a messaging standard for machine generated data*

8. Knowledge Representation

- 8.1. Knowledge representation concerns how knowledge can be represented in a computer-traceable form such that it can assist clinicians and patients making evidence based decision. It is most valuable if the knowledge can be incorporated into the automated systems to support health care delivery at the point of care for improving the quality, safety and effectiveness of clinical care. It includes various areas, e.g. medical literature referencing, reminders and alerts, decision support systems implementation, and guidelines and protocols development.
- 8.2. Knowledge representation is still under development. So far, knowledge representation mainly focus on the development in clinical area, however, there are also needs to incorporate the other elements into the real practice, e.g. legal, social, financial and geo informatics.
- 8.3. Guideline Interchange Format (GLIF) ⁽²³⁾, is developed by the InterMed Collaboratory (comprising Harvard, Stanford, McGill and Columbia universities). It is a computer-interpretable language for modeling and executing clinical practice guidelines. Work is now focusing on developing building-block components that all guideline models must accommodate, such as decision logic expression, how data are to be referenced to, how workflow are to be presented ⁽¹⁾.
- 8.4. Map of Medicine (MOM) ⁽²⁴⁾ is maintained by Informa Healthcare. It works with the NHS Connecting for Health and develops templates on clinical pathways. MOM can be integrated into local systems and supports local adaptation.
- 8.5. The knowledge database on drug will provides information on drug-drug interactions, allergies, contraindications, drug–laboratory inferences, toxicology. They are mainly proprietary products. The most common ones are FirstDatabank, Medi-Span, and Multum.
- 8.6. The Cochrane Library ⁽²⁵⁾ is developed and maintained by the Cochrane Collaboration. The Collaboration analyzes the evidence for and against the effectiveness and appropriateness of certain interventions in a specific clinical situation.
- 8.7. Challenges of knowledge representation ⁽¹⁾ include :
 - 8.7.1. selecting the right knowledge from the exponential increasing medical knowledge

- 8.7.2. guidelines are developed for different purposes, e.g. supporting clinical workflow, facilitate monitoring, driving consultation, and no model fits all purposes
- 8.7.3. lack of comprehensive medical vocabularies,
- 8.7.4. lack of standards to represent knowledge
- 8.7.5. developing comprehensive evidence based knowledge and keeping it abreast of medical advancement
- 8.7.6. translating knowledge into unambiguous computer recognized language
- 8.7.7. delivering sufficient knowledge at the right time without too much disturb to the clinical workflow, and yet still maintaining system performance

8.8. *Recommendations*

- 8.8.1. *To monitor the development of various knowledge representation standards and liaise with the healthcare professional bodies on their suitability to be adopted in Hong Kong.*

9. Data Privacy, Confidentiality & Security

9.1. Introduction

9.1.1. Privacy is the right to control who has access to one's own data, what they can do with those data, and under what circumstances. Confidentiality is the protection of data from inappropriate or unauthorized access or use. Security is the physical protection and preservation of data.

9.1.2. Protecting data privacy provides confidence to a person on who can access to one's own data, and encourage information sharing between a person with one's healthcare provider to facilitate health care delivery. It is essential in health care to ensure the integrity and accuracy of the data. In healthcare, the issue is complicated by one's right in controlling the release to the healthcare provider whom one agreed to.

9.1.3. Information technology increases efficiency of data sharing. There are various convenient portable storage media, e.g. PDA, flashdrive, portable harddisk. Data can be shared via internet or these storage media. Advance technology also introduces threats of leaking confidential health data, whether intentionally or not. Without security measures, health data could be accessed by, disclosed to, or modified by unauthorized persons; or one could hardly identified the person who made such unauthorized access; or the system may not be perform normally due to attack to the infrastructure or applications.

9.1.4. Central to protecting data privacy and security is who can access the data; who has the right to control the access of data; how a person, and the healthcare provider can be identified (please also refer to Chapter 4) and authenticated; how data can be protected from unauthorized access; and how data integrity is maintained in the electronic environment. This relates to developing practice guide on data protection, and employing appropriate information technology to execute and to enforce the data protection principle, and educating the public and healthcare providers on data protection.

9.2. Developing Standard Practice Guides

9.2.1. In overseas, there are various initiatives set out to protect the security of health data, e.g. the Health Insurance Portability and Accountability Act (HIPAA) ^(26, 27) of US, Personal Health Information Protection Act ⁽²⁸⁾ of Canada. The HIPAA also

specifies a series of administrative, technical, and physical security procedures to protect the confidentiality of health information.

9.2.2. In Hong Kong, the protection of personal health care data is governed by the Personal Data (Privacy) Ordinance, and the common law of duty of confidence. The Personal Data (Privacy) Ordinance set out the minimum standard for every organization to follow on protecting ones' data. However, there are special conditions that are specific to healthcare environment in data protection, e.g. disease surveillance, various disease registries, amendment of health data or personal identifier... There should be standard practical guides to address these healthcare specific conditions.

9.3. Applying Technology Standards

Various technology standards have been developed to ensure data can only be accessed or amended by authorized users, e.g. multi-factor authentication, digital signature, e-certificate; data can only be read by authorized users, e.g. cryptography.

9.3.1. Multi-factor Authentication

9.3.1.1. Authentication is the process of determining whether someone or something is, in fact, who or what it is declared to be. The traditional way of using logon-ID plus password, a single-factor authentication, is a weak means of data protection as password can often be stolen, accidentally revealed, or forgotten. More stringent authentication processes such as multi-factor authentication is picking up the trend. There are different types of authentication factor such as:

- ✓ what user knows (e.g. login-ID with password)
- ✓ what user has (e.g. e-Certificate, or token device such as RSA token)
- ✓ what user is (e.g. biometric feature such as fingerprint, retina pattern, face recognition)

9.3.1.2. In the project of Public-private Interface of sharing patient data, HA has also adopted the 2-factor authentication to authenticate the private practitioner who accesses the patient data.

9.3.2. e-Certificate

9.3.2.1. e-Certificate is digital document that vouches for the identity and key ownership of an individual, a computer system or an organization. There are 3 recognized Certificate Authority in Hong Kong namely HKPost, Digisign & HiTrust which are supplying e-certificates which complies to X.509 version 3 standard. HA is currently employing e-certificates from HKPost.

9.3.3. Digital Signature

9.3.3.1. Digital signature applies cryptography (please refer to 9.3.4) to authenticate the identity of a person. It enforces non-repudiation. Once data is amended, the initial attached digital signature will no longer be valid.

9.3.3.2. In Hong Kong, the Electronic Transaction Ordinance admits the legality of digital signature in e-commerce provided that the digital signature is supported by a recognized certificate (please refer to 9.3.2), within a valid period, and used in accordance with the terms of that certificate. However, the legality of using digital signature in health data needs exploration.

9.3.4. Cryptography

9.3.4.1. Cryptography (encryption and decryption) is the science to accomplish confidentiality by encoding information into an unreadable format. It ensures data is shared only between the “key” owners and protects the data from disclosure to unauthorized persons. Cryptography can be applied to a single file, or the whole harddisk. Well known cryptographic algorithms including :

- ✓ Data Encryption Standard (DES), a 56-bit encryption algorithm
- ✓ 3DES, uses DES for 3 times with 2 keys to create an effective 112-bit keyspace
- ✓ Advanced Encryption Standard (AES), supports key sizes of 128 bits, 192 bits, and 256 bits
- ✓ RSA, the most successful public/private (asymmetric) key cryptographic algorithm, typically using 1024-bit keys

9.4. Standards Body on Data Security

9.4.1. There are various data security standards being developed / under development at overseas. The following highlights more well-known ones :

- ✓ ASTM E31.20 on privacy and security (Appendix II)
- ✓ ISO TC215 Working Group 4 on health informatics security
- ✓ CEN / TC251 Working Group 3 on security, safety and quality (Appendix III)

9.4.2. Most of these standards cover the following areas :

- 9.4.2.1. developing policies on protection of health data
- 9.4.2.2. identification and authentication of a person / healthcare provider
- 9.4.2.3. accessing health data
- 9.4.2.4. integrity of health data
- 9.4.2.5. auditing the use of health data
- 9.4.2.6. security of health data in internet environment, data communication

9.4.3. The RSA Laboratories has also developed Public-Key Cryptography Standards (PKCS), a set of standards for public-key cryptography, in cooperation with an informal consortium, originally including Apple, Microsoft, DEC, Lotus, Sun and MIT.

9.4.4. The Internet Engineering Task Force (IETF) ⁽²⁹⁾ focuses on standards and protocols relating to internet architecture and operation such as:

- ✓ PKIX – e-Certificate standard
- ✓ IPSec – encryption and authentication during data transmission
- ✓ S/MIME – secure email standard
- ✓ TLS – encryption of data transmission
- ✓ SSH – secure protocol for remote computer connectivity

9.5. Education on Data Privacy & Security

9.5.1. A general healthcare provider may not have the most up-to-date knowledge on how identified health data can be protected. Protecting health data is a joint effort of all parties who are involved in the system design / development, data management and

data usage. Education on protecting health data includes principles in protecting data privacy, and keeping update on the development of various technical measures on data protection.

9.6. *Recommendations*

9.6.1. *All organizations who will handle health data, whether identified or not, should develop practice guides on how health data are being protected, and make available to her data source, e.g. healthcare seekers, other healthcare organizations.*

9.6.2. *Apart from provision of healthcare, sharing of health data should be kept anonymized as far as possible.*

9.6.3. *To liaise with the IT industry :*

9.6.3.1. *on the ongoing update of the security standards to be adopted for sharing confidential data*

9.6.3.2. *for providing ongoing training on up-to-date information on security of confidential data to the healthcare community.*

10. Management of Data Standards

10.1. Healthcare science is moving forward, technology is changing, and standards are evolving. The above discussions highlight various concerns on standards development. Implementing standards would not be achievable without the help of the IT industry to incorporate the agreed standards to hardware and software to improve data sharing efficiency.

10.2. Implementation of health data standards, in particular on terminology and knowledge representation, is a cultural change to the healthcare industry. Standards compliance may cause a shift to the existing mindset, a change to the current clinical workflow, and an update to the present information systems. Experience indicates the need of modification of the international standards to suit local needs. In addition to issues relating to standards development, other issues need to be considered in standards implementation, e.g. resources for compliance tools, resources for enhancing the existing systems to incorporate standards requirements, buy-in of healthcare industry.

10.3. Recommendations

To establish a Health data Standards Council to coordinate all activities relating to healthcare standards. The council's activity should include and not limit to the following :

10.3.1. *formulate a strategic plan, budget, and timeline on developing and implementing health data standards*

10.3.2. *identify the standards to be adopted in Hong Kong healthcare sector*

10.3.3. *establish a standards development/ratification process*

10.3.4. *identify and prioritize the standards activities to cover critical areas first, e.g. identification, record structure, messaging, terminology*

10.3.5. *develop implementation guide, e.g. related operation principles*

10.3.6. *educate and increase awareness of healthcare standards at various sectors*

10.3.7. *assess standards conformance & certification, including systems developed by the healthcare institute and the vendors*

10.3.8. *audit on standards compliance*

- 10.3.9. seek government's & community support to improve the current patient identification process by :*
- 10.3.9.1. sharing newborn's birth registration number between Immigration Department and the birth hospital*
 - 10.3.9.2. including baby's surname on the Birth Certificate*
 - 10.3.9.3. using biometric data to verify patient's identity*
 - 10.3.9.4. support of simplified Chinese characters for identifying persons from mainland China*
- 10.3.10. liaise with various professional bodies on development of standards that are specific to local environment, e.g. healthcare providers, healthcare institutes...*
- 10.3.11. keep abreast of the international trends on issues relating to health data standards and assess its impact to local environment*
- 10.3.12. liaise with international standards organizations on incorporating local requirements as part of the international standards*

11. Conclusion

11.1. This report only addresses issues relating to standardization of health data. The imbalanced IT development at the public vs the private sector is beyond the scope of the current study. Such disparity provides an opportunity to the adoption of standards when electronic health record is introduced in areas when computerization is limited / not available.

11.2. 千里之行，始於足下。(A journey of a thousand miles begins with a single step.)

The ultimate goal of data standardization and data sharing is to provide a womb-to-tomb record so as to improve both personal health of the Hong Kong residents and public health of the Hong Kong community. The HA is developing Information Architecture by integrating various standards, e.g. record structure, terminology standards, to facilitate sharing of health data. This could be the first step to build the health record for Hong Kong citizens, which could not be achieved without a collaborative effort of the healthcare professionals, IT professionals and health informaticians.

Glossary

3DES	Triple Data Encryption Standard
AES	Advanced Encryption Standard
ANSI	American National Standards Institute
ASTM	American Society for Testing and Materials
BNF	British National Formulary
CAP	College of American Pathologist
CCOW	Clinical Context Object Workgroup
CCR	Continuity of Care Record
CDA	Clinical Document Architecture
CDR	Central Drug Repository
CMS	Clinical Management System
CPT	Current Procedural Terminology
DES	Data Encryption Standard
DH	Department of Health
DICOM	Digital Imaging and Communications in Medicine
ECRI	Emergency Care Research Institute
ePR	Electronic Patient Record
GLIF	Guideline Interchange Format
GMDN	Global Medical Device Nomenclature
HA	Hospital Authority
HACVT	Hospital Authority Clinical Vocabulary Table
HKID	Hong Kong Identity Card
HKPMI	Hong Kong Patient Master Index
HL7	Health Level 7
ICD	International Classification of Diseases
ICD 10 AM	International Classification of Diseases, 10 th revision, Australian Modification
ICD 9 CM	International Classification of Diseases, 9 th revision, Clinical Modification
ICF	International Classification of Functioning, Disability and Health
ICH	International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use
ICPC	International Classification of Primary Care
IEEE	Institute of Electrical and Electronics Engineers
IETF	Internet Engineering Task Force
IFPMA	International Federation of Pharmaceutical Manufacturers and Associations

ISO	International Organization for Standardization
IT	Information Technology
LOINC	Logical Observation Identifiers, Names and Codes
MeSH	Medical Subject Headings
MOM	Map of Medicine
MSSO	Maintenance and Support Services Organization
NCCH	National Centre for Classification in Health
NEMA	National Electrical Manufacturers Association
NLM	National Library of Medicine
OGCIO	Office of Government Chief Information Officer
RIM	Reference Information Model
SNOMED CT	Systematized Nomenclature of Medicine, Clinical Terms
UMDNS	Universal Medical Device Nomenclature System
WHO	World Health Organization
WONCA	World Organization of National Colleges and Academic Associations of General Practice/Family Doctors

References

1. Institute of Medicine, 2004. Patient Safety: Achieving a New Standard for Care.
2. Hong Kong Medical Association, 2006. Report of HKMA IT Survey 2006.
3. Continuity of Care Record <http://www.continuityofcarerecord.org/x6169.xml>
4. Rector, AL. Thesauri and formal classifications : terminologies for people and machines. Mehtods Inf. Med. 1998;37:501-9.
5. Cimino, JJ. Desiderata for controlled medical vocabularies in the twenty-first centry. Methods Inf Med. 1998;37:394-403.
6. British National Formulary <http://www.bnf.org/bnf/>
7. RxNorm <http://www.nlm.nih.gov/research/umls/rxnorm/index.html>
8. MedDRA <http://www.meddrasso.com/MSSOWeb/index.htm>
9. LOINC <http://www.regenstrief.org/loinc/>
10. DICOM <http://medical.nema.org/>
11. ICD <http://www.who.int/classifications/icd/en/>
12. Chute CG, Cohn SP, Campbell KE, Lover DE, Campbell JR. The content coverage of clinical classifications. J Am Med Inform Assoc. 1996;3:224-33.
13. ICPC <http://www.globalfamilydoctor.com/wicc/icpcstory.html>
14. ICF <http://www.who.int/classifications/icf/en/>
15. SNOMED <http://www.snomed.org/snomedct/index.html>
16. CPT <http://www.ama-assn.org/ama/pub/category/3113.html>
17. MeSH <http://www.nlm.nih.gov/mesh/>
18. National Centre for Classification in Health <http://www3.fhs.usyd.edu.au/ncchwww/site/>
19. HL7 <http://www.hl7.org/>
20. Reference Information Model <http://www.hl7.org.au/HL7-V3-Resrcs.htm>
21. Gartner (2006). HL7 v3 messages need a critical midcourse correction
22. IEEE 1073 <http://www.ieee1073.org/standards/1073standards.html>
23. GLIF www.glif.org
24. Map of Medicine <http://www.mapofmedicine.co.uk/>
25. Cochrane Library <http://www.cochrane.org/reviews/clibintro.htm>
26. Health Insurance Portability and Accountability Act <http://www.hhs.gov/ocr/hipaa/>
27. Health Insurance Portability and Accountability Act <http://www.cms.hhs.gov/HIPAAGenInfo/>
28. Personal Health Information Protection Act http://www.e-laws.gov.on.ca/DBLaws/Statutes/English/04p03_e.htm
29. Internet Engineering Task Force <http://www.ietf.org/home.html>

Appendix I

Adoption of Standard Terminologies in Hong Kong Hospitals

		Hospital A	HA hospitals
Patient identification	HKID card number / HK Birth Certificate number	Y	Y
	Self-developed		Y
Diagnosis	ICD 9 CM	Y	Y
	ICD 10	Y	Y
	International Classification for Primary Care v.1	Y	
	International Classification for Primary Care v.2		Y
	International Classification of Functioning, Disability and Health (ICF)		Y
Procedures	ICD9 CM	Y	Y
Laboratory data	SNOMED CT		Y
	SNOMED III		Y
	Laboratory Observation Identifiers, Names and Codes (LOINC)		Y
	Self-developed	Y	Y
Drug data	British National Formulary (BNF)	Y	Y
	USPDI, MartinDale	Y	
Imaging	The Digital Imaging and Communications in Medicine (DICOM)	Y	Y

Appendix II ASTM E31.20 Standards on Privacy and Security

- E1714-00 Standard Guide for Properties of a Universal Healthcare Identifier (UHID)
- E1762-95(2003) Standard Guide for Electronic Authentication of Health Care Information
- E1869-04 Standard Guide for Confidentiality, Privacy, Access, and Data Security Principles for Health Information Including Electronic Health Records
- E1985-98(2005) Standard Guide for User Authentication and Authorization
- E1986-98(2005) Standard Guide for Information Access Privileges to Health Information
- E1987-98 Standard Guide for Individual Rights Regarding Health Information
- E1988-98 Standard Guide for Training of Persons who have Access to Health Information
- E2017-99(2005) Standard Guide for Amendments to Health Information
- E2084-00 Standard Specification for Authentication of Healthcare Information Using Digital Signatures
- E2085-00a Standard Guide on Security Framework for Healthcare Information
- E2086-00 Standard Guide for Internet and Intranet Healthcare Security
- E2147-01 Standard Specification for Audit and Disclosure Logs for Use in Health Information Systems
- E2212-02a Standard Practice for Healthcare Certificate Policy

Appendix III CEN Working Group 3 on Security, Safety and Quality

Secure User Identification for Healthcare - Strong Authentication using microprocessor cards (ENV)

Security for Healthcare Communication (ENV)

Security requirements for intermittently connected devices (ENV)

Safety and Security Related Software Quality Standards for Healthcare (CR)

Framework for formal modeling of healthcare security policies (CR)

Safety procedures for identification of persons and related objects (CR)