# Developing and Using Individual Identifiers for the Provision of Health Services including HIV

## Proceedings from a Workshop, 24–26 February 2009

## Montreux, Switzerland

# Contents

For comments or queries on this report or some of the issues raised in it, please contact Eddy Beck (**becke@unaids.org**) or Xen Santas (**xms1@cdc.gov**)

## Executive Summary

**E.1** UNAIDS, with PEPFAR support, sponsored a workshop on developing guidelines for the development and use of unique individual identifiers for health services, including HIV. A multidisciplinary group of health professionals and people living with HIV were invited to attend the workshop. The health professionals included country program managers, country-based and international information technology experts, clinicians, epidemiologists, program evaluators, ethicists and legal experts.

**E.2** The development and use of unique identifiers will assist countries in the process of developing HIV services within the context of universal access to HIV prevention, treatment, care and support services, in addition to strengthening the country's healthcare system as a whole.

**E.3** It was recognized by meeting participants that services for people living with and affected by HIV include both health and social services, and that to optimize service provision within and across health and social services sectors, strategies and procedures for linking individual service provision records would be needed, including implementation of unique identifiers. After deliberation among all participants, however, it was agreed that the focus of this workshop would be on the development and implementation of unique identifiers for individuals focused on the needs of the health sector, here termed individual identifiers for health services.

**E.4** It was also agreed countries should not develop and implement HIV-specific unique identifiers but they should develop generic, health system–wide unique identifiers. While these identifiers may be implemented for a single vertical, disease-specific program, it is preferable that they should be generated for the healthcare system as a whole. This could improve the continuity of care across the healthcare system as a whole, thereby strengthening the country's healthcare system.

**E.5** The development of healthcare system wide unique identifiers would also reduce the potential for people living or affected with HIV to be stigmatized or discriminated against. It could encourage more people to come forward for HIV testing and counseling, people living with HIV to access preventive and therapeutic services earlier, potentially all resulting in better outcomes of the use of services and enhancing the social and economic aspects of individuals, communities and countries.

**E.6** Unique individual identifiers for health services can help strengthen fragmented health services in countries by linking data held within facilities and enabling the flow of information across the general health system and thereby also enhancing the quality, comprehensiveness and continuity of HIV-specific services.

**E.7** The development and use of these identifiers should be based on the principles specified in the UNAIDS/PEPFAR *Interim Guidelines on Protecting the Confidentiality and Security of HIV Information*. Their implementation will strengthen health information systems overall and thus aid in protecting HIV information confidentiality and security.

**E.8** High-quality information must be collected in order to improve the quality and coordination of service provision through the development of individual longitudinal service records, and to improve the effectiveness, efficiency, equity and acceptability of these services through ongoing monitoring and evaluation.

**E.9** National unique individual identifiers can assist service providers in coordinating services and ensuring that persons receive the full range of necessary services as provided by the country's

healthcare system. They help unify fragmented services systems into a more rationally organized system, and provide policy-makers with critical strategic information that inform strategic planning.

**E.10** Countries from all regions across the world, including high-, middle- and lower-income countries, have developed or are in the process of developing unique identifiers. *Denmark* is a country which has developed national unique identifiers that are used across health and social services in that country. *Botswana, Brazil, Kenya, Malawi, Ukraine, Thailand* and *Zambia* are countries in the process of developing unique identifiers or planning to do so.

**E.11** Like all health information, the development and use of unique identifiers requires balancing the individual's right to privacy and confidentiality with the need for individual-level information to optimize the provision of services to ensure their effectiveness, efficiency, equity and acceptability for both users and providers of those services.

**E.12** Development of unique identifiers should be based to relevant legal frameworks, which should address issues of requirement of consent, how and by whom such information is used, and minimizing the potential risks associated with the use of health information.

**E.13** Health information systems must balance society's interest in accurate information with the rights of the individual to control the use of his or her personal medical information, and to be free from stigma and discrimination.

**E.14** As information on an individual user of health services is often collected and used across multiple facilities, to ensure that records representing the same individual can be identified and linked presents a number of technical challenges. Common approaches to overcoming these challenges has involved the development and implementation of biometric identifiers – like fingerprints, voice scanning, retinal scanning – as well as record matching algorithms based on a range of variables within the patient record.

**E.15** Participants recognized that different countries might adopt different approaches to fit their national circumstances and the nature of their health systems; however, the principles that follow could guide their development in countries.

**E.16** As it is difficult to change national systems once they are in place, countries are advised to take care to make the soundest possible decisions at the outset in putting in place a system for individual identifiers for health services.

**E.17** The intended scope of use for a national unique identifier will determine when an identifier should be assigned. For instance, an unique identifier could be assigned when first visiting a health facility or at birth as part of a vital records registration process.

**E.18** Steps should be taken to ensure that each identifier is truly unique. Personal identifiers for health records ideally should be unchanging, simple, acceptable, easily accessible and available, practical to implement, inexpensive, portable and durable and should ideally contain no information that is derived from the individual.

**E.19** To promote continuity of care, individual identifiers should be readily available for appropriate use in all relevant healthcare settings. Efforts should be taken to ensure that identifiers capture all record fragments linked to a particular individual. These identifiers should be included as one component of a comprehensive portfolio of nationwide standards for health information.

**E.20** Where definitive linkage of a person's health records is not possible using an existing health care identifier, sophisticated matching methods represent one method that may be used to identify the same individual using health services across different locations or time. Such methods can also be used to reduce duplicate records. Various sophisticated strategies for probabilistic matching exist but these need to be evaluated against the data typically contained in the health records of the countries where they might be used.

**E.21** In considering any system, the following issues should be considered: should the unique identifiers be generated peripherally at each clinic or centrally? Are unique identifiers acceptable to the population? Who should assign and supervise the codes, and what procedures should be required? What mechanisms should be in place to avoid errors or duplication?

## 1.0 Introduction

As reflected in the 2006 Political Declaration on HIV/AIDS, unanimously endorsed at the High Level Meeting on HIV/AIDS at the United Nations General Assembly, countries have committed to move toward universal access to HIV prevention, treatment, care and support by 2010. Consistent with this commitment, many countries are in the process of scaling up HIV services. This scale-up is resulting in the collection of an increasing amount of individual-level data. These data are being used to:

 *1) improve the quality and coordination of service provision through the development of individual longitudinal service records, and*
*2) improve the effectiveness, efficiency, equity and acceptability of these services through ongoing monitoring and evaluation.*

Use of an unique individual identifier for health services including HIV, for each individual will advance these aims. Unique identifiers enable all data collected within a facility to be correctly attributed to a specific person. In addition, where persons receive services from a number of different facilities, relevant information can be more effectively and efficiently shared and linked across service sites to improve service coordination and strengthen monitoring and evaluation. The continuing prevalence of HIV-related stigma and discrimination in diverse settings underscores the imperative to protect the confidentiality and security of all personal data.

To address the issue of protecting HIV-related data, a workshop was held in Geneva, May 2006, which produced the *Interim UNAIDS/PEPFAR Guidelines on Protecting the Confidentiality and Security of HIV Information* [1] ("*Interim Guidelines*"). These guidelines provide guidance to middle- and lower-income countries to develop, adopt and implement culturally appropriate measures to protect personal data.

While protection from harm is a central consideration in any data collection scheme, the rules, regulations, procedures and practices governing the collection of health data also affect the degree to which such personal information is available to improve patient health and medical outcomes. In devising national frameworks for the collection and use of personally identifiable patient information, countries must balance interests associated with:

a) *maximizing benefits that can and should come from the wise and fullest use of data;* and
b) *minimizing the harm that can result from either malicious or inadvertent inappropriate release of individually identifiable data.*

These potential benefits and harms may accrue to individuals, groups, or institutions. Several factors motivated the development of a set of principles or guidelines that may help maintain this balance. These factors are independent of context and include both the tremendous potential of patient health data repositories and the potential for increased risk of confidentiality breach inherent in consolidated and centrally accessible data.

The *Interim Guidelines* deal with three interrelated concepts of *privacy*, *confidentiality* and *security*, and all affect the protection of sensitive data. *Privacy* is both a legal and an ethical concept. Legal aspects of privacy pertain to formal protections accorded to an individual to control both access to and use of personal information. Privacy protections provide the overall framework within which both confidentiality and security are implemented. *Confidentiality* relates to the right of individuals

to have their data protected during collection, storage, transfer and use, in order to prevent unauthorized disclosure of that information to third parties. *Security* is a collection of technical approaches that address physical, electronic and procedural aspects of protecting information collected as part of the scale-up of HIV services. Security must address both protection of data from inadvertent or malicious inappropriate disclosure and non-availability of data due to system failure and user errors.

To determine the degree to which countries had developed or implemented such guidelines, UNAIDS and PEPFAR constructed a questionnaire based on the *Interim Guidelines*. Of the 78 countries that responded to this questionnaire, only 21 had developed some sort of guidelines. However, none of these guidelines covered all of the areas recommended by the *Interim Guidelines* [2].

## 2.0 Individual Identifiers for the Provision of Health Services Including HIV

A major impediment to the scaling up of HIV services is the relative weakness and fragmented nature of healthcare and social service systems in many middle- and lower-income countries. The use of unique identifiers could help mitigate the effects of fragmented health systems. Many people living with HIV receive health services in multiple settings, such as antiretroviral clinics, TB clinics and antenatal settings. Many also change providers, often because they have moved to find work. As in the therapeutic context, HIV prevention services are also often delivered in multiple settings, including various health sector or non-health-sector settings.

By linking patient data horizontally across the health system, an unified system of unique identifiers assists providers in coordinating services and ensuring that persons obtain the full range of necessary services. Individual identifiers for health services help strengthen health systems, linking disease-specific, or vertical, systems into a comprehensive, rationally organized system of health care delivery.

For those individuals using prevention or treatment services, linked data for individual patients can help clinicians improve services, including avoiding the prescription of medications that have negative interactions with drugs prescribed at another facility. Information-sharing is important for a variety of different settings including patients who may be receiving both HIV and TB drugs and pregnant women living with HIV who need to switch post-partum from antenatal care to long-term HIV care.

In the absence of systems that track the use, cost, outcome and impact of services for persons or groups of patients, policy-makers and clinicians will continue to develop and implement strategies that are not fully informed by critical information.
As follow-up to the *Interim Guidelines*, with PEPFAR support, UNAIDS sponsored a workshop on developing guidelines for the use and development of unique individual identifiers for health services. The participants included relevant health care professionals from eight countries involved with developing their HIV information infrastructure[1], people living with HIV, information technology experts, ethicists and public health professionals (Appendix 1). This report summarizes the methods used, main discussions, conclusions and recommendations from this workshop.

---

[1] Countries whose experiences with health services identifiers were discussed during the meeting included Botswana, Brazil, Denmark, Kenya, Malawi, Thailand, Ukraine and Zambia.

## 3.0 Methods

Prior to the workshop, all participants received an annotated bibliography on unique identifiers produced by the U.S. Centers for Disease Control and Prevention and MACRO International. This bibliography provided the participants with brief summaries of selected literature regarding unique identifiers (Appendix 3).

The three-day workshop included introductory presentations on ethical and technical aspects of unique identifiers, while representatives of the eight different countries presented their experiences in developing HIV information systems and existing unique identifier schemas or those under development. Following these plenary presentations, participants assembled in five different sub-groups to examine various aspects of unique identifiers in the context of HIV services, reassembling in plenary to examine and discuss the various sub-group findings and to discuss and agree on recommendations. In assigning participants to individual working groups, efforts were made to distribute key areas of expertise (e.g., technical, legal, ethical, ministerial, civil society) across the different groups.

Each working group was assigned a specific topic to consider in developing its recommendations. The focus of each working group was as follows:

*Working Group 1*: information collected for provision of clinical or social services from a single system within the same clinic, but collected and stored over time.

*Working Group 2*: information collected for provision of clinical or social services, but collected from different systems either within the same facility or from geographically separated units. This could include data collected from separate laboratory information systems or pharmacy databases, as well as from clinics providing treatment services at the community level.

*Working Group 3***:** information collected for provision of clinical and other care, from clients who receive services from multiple providers. Examples include people who have been tested and whose management moves from a counseling and testing clinic to an antiretroviral therapy (ART) or other care/treatment clinic, women diagnosed to be infected with HIV while using antenatal care (ANC) services and who switch to an ART clinic after giving birth, or persons who move from inpatient to outpatient services within the same health facility or between health facilities.

*Working Group 4:* information collected for monitoring services. This includes data collected at district, regional or national levels to monitor program progress and quality and to assist in de-duplicating counts or identifying cases lost to follow-up.
*Working Group 5:* information collected for program evaluation or for research. This includes data, including anonymized or pseudo-anonymized individual-level data, collected at district, regional or national levels and stored in data warehouses or chronic disease registries.

Each working group was asked to address four specific aspects within their deliberations:

(1) *design and implementation of identifiers, with particular focus on format and technologies;*
(2) *legal, ethical and cultural considerations associated with unique identifiers;*
(3) *use of national level identifiers and their relationship to individual identifiers used for health services;* and
(4) *use of probabilistic matching algorithms.*

Following reports from each of the working groups, a drafting group met to develop a consolidated set of recommendations for moving forward as countries consider implementing unique identifiers for health services. Meeting in plenary, participants reviewed, discussed and revised these consolidated recommendations.

## 4.0 The Use and Development of Unique Identifiers for HIV Services

During plenary discussions, workshop participants addressed the focus of the guidelines. It was agreed that unique identifiers should pertain not only to health care services but also for enabling social services that promote the effectiveness and continuity of health services. After some discussion, participants agreed that the guidelines being developed as part of the workshop would primarily focus on unique identifiers for the healthcare sector, with the expectation that such identifiers could ultimately be extended to, or integrated with, unique identifiers used in other sectors. Some participants urged caution, however, in linking health services identifiers with more general-purpose identifiers, because of the potential increased risk for breach of confidentiality and stigma.

Participants also extensively discussed whether countries should develop HIV-specific identifiers or generic identifiers applicable to both HIV-related and non-HIV-related health services. Given the stigma and discrimination associated with HIV and the fact that unique identifiers can be used across a broad range of health services, it was recommended that development of generic unique identifiers to be used across the health sector was the most sensible approach.

## 5.0 Country Experiences with Developing Unique Identifiers

Representatives presented experiences from a number of middle- and lower-income countries that are involved in developing their HIV-information systems. In addition, one participant described the health information system in Denmark, an industrialized country that has a well-developed and long-standing national health and HIV information system using unique identifiers.

### 5.1 Denmark

Denmark is a high-income country of 5.5 million people. It has a well-developed and organized national administrative infrastructure and high population literacy.
Beginning in 1968, Denmark assigned an unique identifier to all Danish citizens. Information from this system is stored in a Central Person Registry (CPR), and each citizen is assigned an unique CPR number. This number is used for vital registration (births and deaths), citizenship, immigration or emigration and linking families. The CPR number enables a number of other Danish registries to be linked, including the Danish hospital registry of inpatient and outpatient information, the death registry, the national cancer registry and a social registry that collects socio-economic information as well as prison information. The 10-digit CPR number includes 6 numbers based on the person's date of birth, plus an additional 4 numbers, some of which are derived from demographic characteristics of the individual.

HIV services in Denmark are publicly financed and provided free at point of service delivery, with treatment services provided at eight centers distributed throughout the country. All people living with HIV managed at these eight centers are provided with a specific unique HIV number, which is different from the CPR number but linked to it. Linkage between these databases permits individual

HIV-related data to be linked with other health registries including hospital registries, vital statistics or cancer registries and relevant socio-demographic data. HIV infection is monitored in Denmark through The Danish HIV Cohort Study uses an unique identifier which is different from the CPR number, and links HIV healthcare information including demographic, tests results, treatment and outcome information. As a link exists between the HIV unique identifier and the CPR number, data from the health and social services registries can be linked with HIV data. This enables analyses to be made on mortality and morbidity patterns, impact of interventions or services on utilization patterns and outcomes [3] and comparisons of service utilization and medical outcomes between HIV-infected and HIVuninfected Danes [3,4]. The Danish HIV Cohort Study is governed by a Steering Committee, which includes one member from each of the eight HIV treatment centers in Denmark.

Denmark's health information system has permitted authorities to track rates of HIV-related hospitalization over time and to identify factors that appear to increase mortality in the general population, people living with HIV and people with other illnesses. Such information is not only useful for service providers, but it also facilitates population-level monitoring of health trends, thereby providing important information for health professionals, policymakers and other stakeholders [5].

### 5.2 Botswana

Botswana is a land-locked country in southern Africa with a population of about 1.8 million. Botswana is one of the countries with a high HIV prevalence with a population prevalence of 17.6%. It is a middle-income country with a commodity-driven economy. The health system is decentralized and distributed across 24 health districts. The country has two main referral hospitals in Gaborone, the capital city, and in Francistown, the second major town in the country. There are also District hospitals in all the districts, including clinics and health posts. Most of the clinics and hospitals provide comprehensive services, which include ARV therapy.

The country adheres to the UNAIDS "Three-Ones" Principles, meaning it has one National AIDS Coordinating agency (NACA), one National Strategic Framework (the first one covered the period 2003–2009 and the second is currently being developed) and one Monitoring and Evaluation system, called the BHRIMS, or the Botswana HIV/AIDS Rapid Information Management System. Botswana reports the HIV UNGASS indicators to the UNAIDS Secretariat every two years. The strong political commitment by the government contributed to a comprehensive Botswana HIV National response. In addition to scaling up comprehensive treatment and care, including counseling and testing services, prevention of mother-to-child transmission services and antiretroviral treatment, extensive prevention services were also developed, which currently include roll-out of the "Safe Male Circumcision Programme" as an additional prevention strategy (Table 1). Citizens of Botswana are provided with health services which are almost free of charge at point of delivery, though clients pay a small user fee of BW Pula 2.00, which is about US$0.35. ARV therapy is free at point of services delivery for all citizens.

In Botswana, the National Identification document is called the "Omang ID card." This card has a number, which is used for all purposes of registration, identification and other social services, including for health care in those hospitals that have the Integrated Patient Management System (IPMS). This number will also be used in all the ARV clinics where Patient Integrated Management System II (PIMS II) will be installed, which is currently being rolled out nationwide.

The Omang number consists of a series of different numbers, without any letters. Initially assigned only to individuals who reached age 16, the Omang number was later subsequently expanded to include all live births. The Omang number is issued and maintained centrally. As the number is unique for each individual, it has the potential to identify and track a person across multiple facilities, programmes or services and would allow for the linkage of information for a lifelong view of a person's medical and social history.

**Table 1: Different Government Programmes Part of Botswana's HIV Response**

| Late 1980s | Information Education and Communication (IEC) |
| --- | --- |
| | Screening of blood and blood products |
| 1990 | AIDS at the Workplace Programme |
| Early 1990s | Sexually Transmitted Infections (STI) Control Programme |
| 1995 | Community Home-Based Care (CHBC) |
| 1999 | Prevention of Mother-to-Child Transmission (PMTCT), rolled out nationwide by 2001 |
| | National AIDS Coordinating Agency established |
| 2000 | Isoniazid Preventative Therapy (IPT) |
| | Voluntary Counselling and Testing (VCT) (Tebelopele), |
| 2001 | MASA Program to provide antiretroviral therapy |
| 2004 | Routine HIV (Opt-out) Testing Programme (RHT) |
| 2009 | Safe Male Circumcision Programme |

As an example of the uses of the Omang number, the Tebelopele voluntary counselling and testing centres started using the Omang number in January 2007. Adoption of the Omang number by Tebelopele has permitted a more accurate picture of the number of people tested while avoiding double counting, as well as possibly estimating the rate at which individuals that previously tested HIV-negative have seroconverted. The unique identifier has also facilitated contact tracing and follow-up of people living with HIV referred to other services, including antiretroviral or hospital services. Foreign nationals use their passport number for health services registration.

While a number of programs now use the Omang number, some also still use site-specific identification numbers or use the Omang in combination with a specific program identifier. For instance, the antiretroviral services in the country's four tertiary centres have their own identification number called the ARV number. It consists of the letters "ARV" followed by a number consisting of several digits.

Problems can arise through lack of consistency at registration. For example, a number can be recorded in a variety of ways such as "ARV-25476397," or "ARV25476397," or "ARV 25476397." The space or dash inserted between the letters "ARV" and the numbers needs to be consistent to allow identification of the same individual. These problems are being addressed through probabilistic matching being done on regular basis before data entered into the ARV data warehouse, after being collected and compiled from all ARV sites using different identifiers. District hospitals use another site-specific number. Tuberculosis services use the Omang combined with another identifier. Consequently, identifiers may be unique only at the issuing site, and individual patients may sometimes have more than one identifier at a particular site or for a specific programme or service.

Other problems identified to date include occasional incorrect entry of the Omang, resulting in nonexistent or duplicate numbers, their potential misuse through sharing the number for one

individual with another individual , using somebody else's number or the number simply not used or entered correctly. In situations where the Omang is not used as the sole and unique identifier, it may be very difficult to link patient records or to avoid duplicate records.



Plan for Integrating Health Systems in Botswana
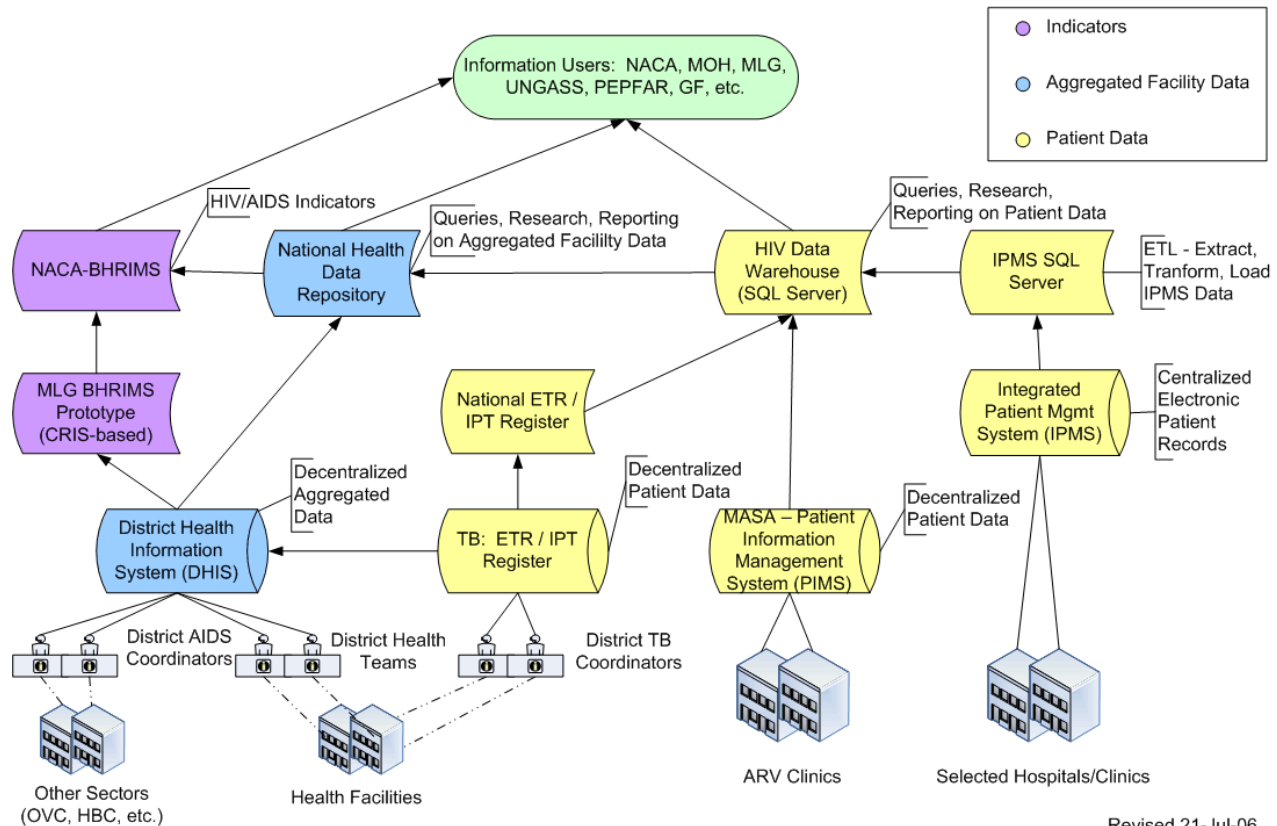
Though there is strong collaboration between different national and international actors, some other challenges facing Botswana health information systems involve the inadequate networking between different health information systems in the country, the insufficient harmonization of data and data format, and human and technical resource issues, including information technology support, equipment and training of staff. Another challenge that will need to be addressed is to ensure the confidentiality and security of the stored health data at the various sites or in data repositories. Addressing these issues will require the development and implementation of a national health management information system (HMIS – Figure 1), the development of a national HMIS strategic plan, the harmonization of data flow while ensuring different data security access levels, and integrating the different systems within the Ministry of Health.

### 5.3 Brazil

Brazil is a massive country of 187 million people that includes the tropical North, Central Amazonia and the more temperate South and South- East. Brazil's 27 states include 5,564 municipalities, with governmental functions divided among federal, state and municipal levels. Each level of government has a role in developing and implementing health policy, with extensive intergovernmental collaboration required for coherent policy development.
Seventy-nine percent of the population uses the national health system, the *Sistema Unico de Saude* (SUS). Access to services is free at point of delivery. Among more than 600,000 people in Brazil living with HIV, around 190,000 are currently receiving antiretroviral therapy, which has been free of charge since 1996.

Four main health information systems currently exist in Brazil:
- National Notification System: records notifiable diseases including AIDS
- National Health Establishment Database
- National User Database
- National STD and AIDS Department
  - SICLOM – records medications
  - SISCEL – records laboratory tests and results
  - SISGENO – genotyping

Health facilities in Brazil range from large hospitals to small private clinics or medical offices. The National Health Establishment Database obtains information on these diverse public and private establishments. Information is collected from out- and inpatient care, laboratory services, human resources and physical resources, including beds and equipment. Each establishment forwards the required information to municipalities that are responsible for verifying and forwarding it to the health ministry. Roughly 189,000 medical establishments and more than 2 million health professionals are currently registered in the country's health database. All public health facilities must use this database in order to receive payments for services provided.

Brazil currently has no national unique identifier for health services. However, all professionals employed or seconded by SUS have their own unique identifier. Similarly, patients with complex diseases and some who receive services as part of national programs, such as renal replacement therapy and pregnancy programs, have been given unique program-specific identifiers. Currently, about 54% of the Brazilian population has an unique identifier.
Although the National User Database contains approximately 140 million records, the database includes many duplicate records and it is estimated that approximately 100 million people are actually captured by the National User Database. Probabilistic matching is used for epidemiological monitoring, which focuses on AIDS cases only, as HIV is not reportable in Brazil.

Because of their autonomy, municipalities develop their own health information systems. The integration of the different systems is currently lacking in Brazil, but the government has plans to integrate the various systems. One integration strategy is the development and implementation of a National Health Card, which would be based on an unique identifier for each Brazilian.

### 5.4 Kenya

A country of 36 million people, Kenya has an adult HIV prevalence rate of 7%, representing an estimated 1.4 million people living with HIV. HIV prevalence is higher in urban areas than in rural settings.

Kenya issues national unique identity card numbers to all persons older than 18 years. The identity card is used in a variety of labour-related issues. A separate unique identifier has been created for the country's antiretroviral treatment program, which is not linked to the national identity number. The unique identifier number for antiretroviral treatment was designed to maintain an accurate number of people on antiretroviral drugs, facilitate patient tracking within and across facilities, monitor transfers in or out of particular facilities and identify those who are lost to follow-up or who have died.

The unique antiretroviral code contains 11 digits including a 3-digit number for the district, a 3- or 4-digit code for the facility, and 4- or 5-digit code for the individual patient. The numbers assigned to specific districts or facilities have changed in the past, generating potential duplication. Similarly, if patients move and change facilities, they will get a new facility number. Linking data is often difficult or impossible because multiple lists of changing health facility codes exist and different programs use different codes. However, health facility codes are in the process of being standardized.

Unique health facility identifiers would permit the assignment of an unique number to each patient, which they would keep even if they used services at other sites and which could ensure harmonization across facilities. In considering such a system, a number of issues should be considered: should the unique identifiers be generated peripherally at each clinic or centrally? Are unique identifiers acceptable to the population? Who should assign and supervise the codes, and what procedures should be required? What mechanisms should be in place to avoid errors or duplication?

*5.5 Malawi*

In this country of 12.5 million people 12% of the adult population are infected with HIV, giving rise to 85,000 AIDS deaths per year. In 2008, 1.4 million HIV tests were performed in 650 site-specific clinics and 250 outreach programs. Of the 650,000 women who become pregnant each year, 93% are seen for antenatal care in at least one of 560 antenatal sites and 58% of women deliver in a health facility. Antiretroviral treatment is delivered in 221 clinics throughout the country. Of the 223,500 people living with HIV registered for antiretroviral treatment, 147,500 were known to be alive by the end of 2008. Other HIV-related services include support for orphans and nutritional support for certain populations.

Less than 10 percent of the population has been issued a passport for travelling abroad. A national identifier is in the process of being introduced in Malawi, with a phased implementation that currently involves 14 of 28 districts. The first phase focuses on registration of individuals through village registers. The second phase involves registering all adults through a centrally controlled identification system, networking all district registration offices and expanding the issuance of birth certificates.

In order to access health services, all adults and children are provided with a patient health passport. The health passport is mandatory for access to all health services and records all diagnoses and interventions, but it does not include an unique identifier. Similarly, a variety of paper- or electronic-based medical records systems exist for HIV service provision, but all employ system-specific identifiers. As a result of this heterogeneous system, different patients possess different health passports. Due to inadequate privacy protections and the risk of discrimination, health workers living with HIV may seek care at other clinics or provide false information. The paper registers often bear the name of the patient, which compromises the confidentiality and security of the data.
Health care professionals and other stakeholders are now developing a national health system information framework. In 2008 a Data Standards Task Force was formed to harmonize indicators, develop a data warehouse at the Ministry of Health and investigate and develop a system of national identifiers for health services. National standards for health data have been discussed and will be further addressed by the Task Force.

*5.6 Ukraine*

Ukraine has the worst HIV epidemic in Europe. Out of a population of 46 million people, 91,267 people living with HIV were registered by October 2009, of whom 11,506 had AIDS and 14,256 were receiving ART. By the end of 2008, 1.3% of the total adult population were estimated to be living with HIV. Among groups of injecting drug users (IDUs), HIV prevalence ranges from 18% to 63%.

Diverse HIV-related services are available in Ukraine, including prevention services for most-at-risk populations, voluntary counseling and testing, methadone substitution therapy, post-exposure prophylaxis, prevention of mother-to-child transmission, and comprehensive treatment and care services, including services for co-infected persons and sexually transmitted infection management. Data that describe service provision are collected through a variety of different mechanisms. The *EpiAIDS* system captures routine HIV surveillance, recording all registered HIV cases, mode of transmission, AIDS diagnoses and AIDS deaths. Staff at regional AIDS centers enter data for this system, using an unique 16-digit generic client code based on birth date and name. Analyses from this system are based on aggregate data and limited to the national level analyses.

A clinical and antiretroviral monitoring system tracks people in need of treatment, CD4 counts, viral load, clinical staging, number of patients on antiretroviral regimens, HIV/TB patients, people with Hepatitis C, opportunistic infections, treatment regimens, treatment results and other analyses. Data are updated at each patient visit. At the facility level, staff collect personal information, including full name and date of birth. Individual-level data obtained for clinical tracking of individuals is not available for monitoring and evaluation at the national level, as national and international indicators are based on aggregate data. An electronic clinical database is currently in the final stage of development.

The *SyrEx* database primarily collects data from HIV prevention projects among injecting drug users, sex workers, men who have sex with men (MSM) and prisoners. The system counts unique individuals, rather than episodes of service delivery. A coding system permits monitoring of individual service utilization. NGOs working in the prevention field use different coding systems, and an unique identifier code is currently being piloted in one of the regions of Ukraine. If successful it could be implemented throughout the country. A client must receive a minimum package of services in order to be classified as having received the appropriate care, which for IDUs includes the provision of syringes, condoms, information material, consultation with a social worker and referral to other specialists if required.

Ukraine has confronted a number of challenges in its efforts to devise an HIV information system. These include the existence of multiple, non-integrated service delivery networks and data collection tools, unlinked data collection systems and insufficient funding and government leadership. To date, the Ministry of Health has issued no guidance or standards on unique identifiers. However, the strong collaboration that exists among clinics and NGOs on monitoring and evaluation provides a strong basis for piloting the use of unique identifiers within HIV services. Such a pilot project would also benefit from the substantial overlap among users of prevention, treatment, care and support services.

### 5.7 Thailand

Thailand initiated a pilot antiretroviral treatment program in 2000 and extended it in 2006, when antiretroviral treatment became free at the point of service delivery, as part of the National AIDS

Program. Thailand has an extensive and linked health care system, ranging from community hospitals, provincial-level general hospitals and regional-level health facilities.

At the registration of a birth, the Ministry of Interior (MOI) issues all Thai citizens a 13-digit personal identification number (PID). The registration of a birth needs to be completed within 15 days after birth and the first digit of the PID refers to nationality, whereas the $2^{nd}$ to $5^{th}$ digits refer to province and district codes of permanent address, the $6^{th}$ to $12^{th}$ digits is the birth registration number and $13^{th}$ digit is a verification code or check digit. The MOI network collects all births and deaths at the district level and the system is designed to work on a real-time basis.

All health facilities record data on treatment and care for people living with HIV, with patient data stored in a central disease management information system (DMIS). DMIS is a web-based application with a centralized database maintained at the National AIDS Program (NAP). Main objectives are to serve management and reimbursement purposes by the National Health Security Office (NHSO). The PID permits identification of patient-level information and is the central identifier in this system. Collected at the first time of registration, the PID number is then encrypted and stored in the central database. The database allows only the hospital user to retrieve their health data from the database. The encrypted number PID becomes the National AIDS Program (NAP) number, facilitating the linkage of HIV-related data with other patient health information. No one but the patient knows his or her own PID.



Centralized data networking
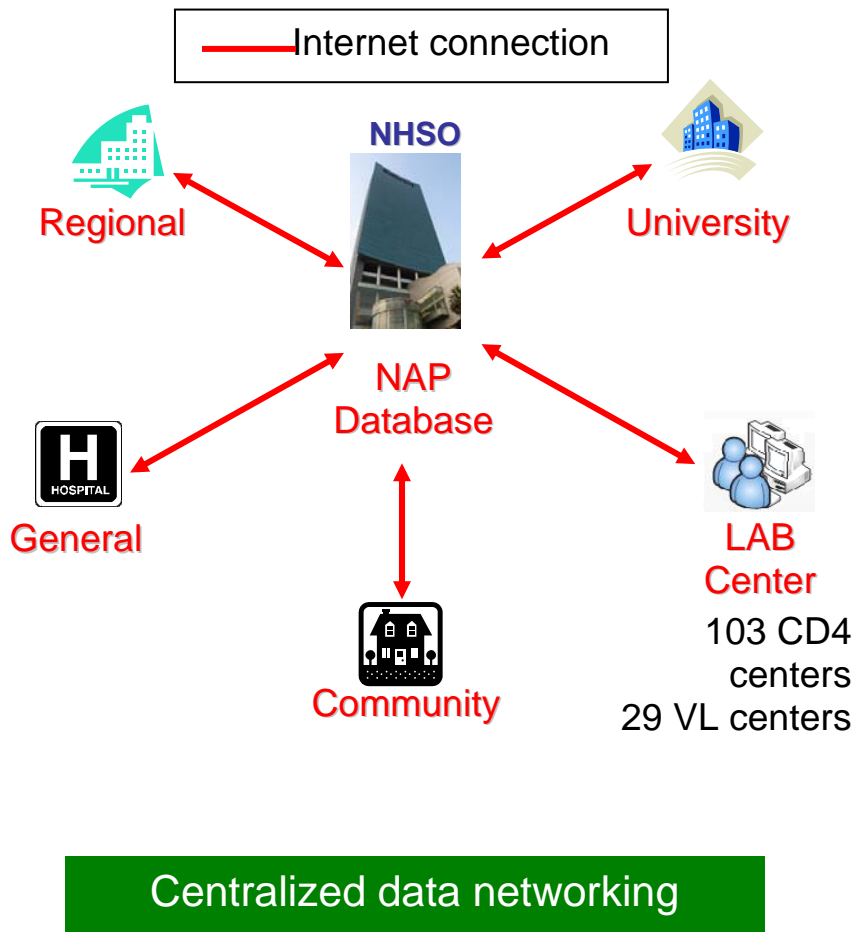
The National AIDS Program database consists of four core modules and four additional modules. The core modules include registration, follow-up, authorization of second line antiretroviral regimen, laboratory requests and reporting. The additional modules collect data from voluntary counseling and testing services, prevention of mother-to-child transmission, administration of post-exposure prophylaxis and monitoring and evaluation. The NAP number permits linkage of data from each module. Through the NAP number, data can link to many external data sources for logistic purposes, assess the quality of the data and monitoring and evaluation of services. Only incoming data from MOI, such as birth or death registrations, are individual records that include PID. All outgoing data from NHSO are individual records without the PID but with the NAP number. For instance, mortality data are transferred from the MOI to the NHSO on a daily basis. The PID from mortality data is then programmatically matched with the PID in the NAP database. NHSO also keeps records of individual inpatients. By matching on the PID number, morbidity patterns of PLHIV can be monitored. Thailand is going to use the PID as an universal identifier for both health and non-health services.



## 5.8 Zambia

Zambia is implementing an electronic health record system – *SmartCare* – employing touch-screen and smart card technology to aid data capture and to facilitate personal portability and control of confidential electronic records. This national program aims to provide every Zambian with a life-long electronic health record (EHR) that assures continuity of high-quality and confidential care, by providing timely information to caregivers at every point of service and to health policy makers through integration with the national health information system.

The system was first proposed in late 2003 as a solution to address the needs of clinics both on and off the internet, and was funded as a pilot in late 2004. Consensus was reached in 2005 for using a locally generated but nationally unique identifier, which could be implemented at clinics using manual or computer processes. The electronic health record assigns global unique identifiers

(GUIDs) for internal management of individual records. By early 2009 the program provided electronic health records to more than 250,000 people. Over 300 district and provincial health leaders have been trained in the use of the electronic health record system, working in over 400 facilities equipped with health record systems workstations.

There are currently five versions of *SmartCare* based on the same infrastructure, including the same shared software code, that provide solutions for developing and maintaining longitudinal and continuous personal service records in the absence of reliable networks. In addition to the national EHR system,

- Zambia has a *SmartDonor* system for the Zambian National Blood Transfusion Service's healthy donor tracking and "loyalty program" and the Zambia National Passport and Citizenship Office is piloting *SmartReg*, a new component of the vital registrations system,
- Ethiopia has adapted *SmartCare* to Ethiopian MOH standards, implemented the EHR fully in the Dire Dawa region and is extending into an adjacent region, as well as piloting other locations.
- Eastern Cape, South Africa has adapted *SmartCare* to satisfy Eastern Cape MOH requirements and has begun piloting the paper fail-over system while developers are completing the clinical software modules to match the new forms.

An EHR system that can use portable smart cards as a record option offers important benefits in the disconnected Africa context, enhancing the ability to link patients between physically separate and perhaps diverse services. This is difficult to do with paper systems in clinics that lack phones, faxes, copiers and other basic paper system processing equipment. Such linkages promote continuity of care, facilitate delivery of appropriate services for basic preventive services such as childhood immunizations and more medically complex conditions such as HIV, where knowledge of previous care is required to properly deliver services. The second important role of Zambia's EHR system is to contribute data for monitoring and evaluation of health services and to help generate sustainable surveillance data for prevention, treatment, care and support services.

The system incorporates coding, compression and encryption features to ensure the confidentiality and security of the data stored on the card. Facility-level data based on individual records are merged across all facilities at the district health office for linkage and de-duplication of counts but are not individually accessible above the level of a facility visited by the client. They are then de-identified and merged at provincial and national levels for program management, epidemiology and other MOH purposes.

## 6.0 Key Points in the Development and Implementation of Unique Identifiers

A host of considerations come into play as countries examine possible options for individual identifiers for health services, including the practical, technical and financial challenges of implementing confidentiality and security, including ensuring the patient's informed consent for the collection and use of personal information. Careful planning is needed to clarify the precise ways that health information is to be used and to ensure that improved personally identifiable information systems contribute to more favorable medical outcomes for people living with HIV or other conditions. Developing and implementing individual identifiers for health services should occur as part of a broader effort to strengthen national data systems generally.

Including affected communities is also critical to building strong, broad-based support for unique identifiers. Affected community stakeholders include non-governmental organizations, community-

based organizations, faith-based organizations and representatives of sexual minorities. An additional systematic description of items that need to be considered and addressed for developing unique identifiers can be found in Appendix 2.

## 6.1 Ethical considerations

Ethics in medicine requires an appropriate balance between broader public health aims and the rights of the individual. In the case of personally identifiable information, this requires a balancing of the individual's rights to privacy and confidentiality with the need for efficient and effective access to these records to promote individual and public health goals.
The ethical principle of respect for persons may require the patient's consent for the collection, distribution and use of personally identifiable data. Throughout the world, various models of consent are reflected in national legal frameworks:

- *Full explicit consent* – This approach requires the patient's clearly expressed consent each time personally identifiable information is to be used. This approach is the most respectful of individual autonomy, but it is also costly and burdensome for providers and patients. Full explicit consent is normally required in the course of medical research, but is not usually demanded for routine care or public health surveillance.
- *Opt-in* – The patient provides explicit consent during the initial contact and "opts in" to allow additional, appropriate use of personal information. The patient reserves the right to "opt out" – or prevent the further use of information – at any point. This approach favors autonomous decision-making, but it may result in incomplete information as some patients may elect not to opt in.
- *Opt-out* – Consent is assumed unless the patient explicitly opts out. This approach is respectful of individual autonomy but, like opt-in, may also result in incomplete information as a result of patients who opt out.
- *No consent* – Under this approach, information is held in trust with assurances of confidentiality. Even where individual consent is not obtained, individuals should nevertheless be notified or informed that their personal data may be shared with others, but only in a format where the data are no longer personally identifiable.

The 2007 *UNAIDS/PEPFAR Interim Guidelines* recommend that relevant cultural norms should be taken into account in deciding which particular approach to informed consent is used in a particular context.

## 6.2 Technical considerations

As information on a single patient is often collected and used across fragmented settings or systems, ensuring that records representing the same individual or entity can be identified, linked and meaningfully analyzed presents a number of technical challenges. For example, there are numerous impediments to accurate matching of individual patients with their relevant personal information, including recording errors, changes in identifiers such as a change of residence or a name change following marriage, the sharing of particular identifiers by multiple individuals, and data quality issues, such as incomplete and non-standardized data fields.

Ideally, an identifier for an individual patient should be
- unique;
- ubiquitous and available for every person;

- unchanging, for instance eye color, date of birth, genotype;
- uncomplicated and easy to recall or record;
- non-controversial, not involving the collection or dissemination of sensitive data, and
- easily and inexpensively accessible.

Potential approaches to uniquely identifying patients include:
- *Biometric identifiers* such as fingerprints, voice scanning, retinal scanning or iris scanning. The advantage of biometric identifiers is that they tend to be highly specific to each individual, but they require additional hardware, may raise privacy and security concerns, and are subject to false non-matches due to hardware or procedural failures.
- *Patient matching algorithms,* which combine multiple patient attributes, such as date of birth, mother's maiden name and other common identifiers, to create an individual identifier. This approach leverages available information and does not require expensive hardware or the generation of an unique identifier that a patient may lose or forget. However, such algorithms are not foolproof due to the possibility of duplication or false positives, and their accuracy varies with data quality. Identifiers comprised of a string of characters require a mathematical scheme to permit validation of identifiers for the purpose of matching patients with specific records. Probabilistic matching algorithms, ranging from the simple to the complex, may be used to help determine whether a particular set of records can be appropriately linked to a specific patient. More precise algorithms require more computing power and may be more suitable for retrospective analysis than for real-time clinical care.

The specific scheme used to generate an unique identifier will inevitably have consequences for health policy and will also inevitably involve the balancing of competing considerations. For example, building an identifier on an easily remembered piece of information, such as birth date or mother's maiden name, could simplify the location of medical records if a patient forgets his or her card during a clinic visit; but the simplicity of this approach could also facilitate fraud or loss of privacy, because those who possess certain basic facts about an individual might more easily obtain his or her personal data.

The desired longevity of a patient identifier affects the design of the identifier schema and is an important input for developing an overall identifier strategy. Specifically, if identifiers are to last more than 100 years, i.e., the maximum lifespan for most individuals, it is necessary for the identifier to contain more than nine characters. The increased identifier length, and thus increased longevity, must be balanced against the fact that longer identifiers increase complexity, raise the likelihood of recording errors and reduce patients' ability to remember their identifiers. Experience suggests that dividing longer identifiers into shorter blocks, such as three sets of four characters for 12-character identifiers, mitigates the disadvantages of longer identifiers. Likewise, health systems that are beginning to build identifier systems should ideally choose identifier schema that will allow for incremental changes over time, such as the conversion of paper records into electronic systems.

## 7.0 Constructing and Implementing Unique Identifiers for Individual Health Records

The meeting participants identified consensus recommendations for unique individual identifiers for health services, which are summarized below. These recommendations aim to identify key principles and best practices for identifier systems, recognizing that diverse countries may and should adopt various approaches to fit their national circumstances and the nature of their health systems.

In order to uniquely identify all current individuals within a country and accommodate future additional individuals, the length of unique identifiers can exceed ten digits.  As identifier length increases, so too does the probability that typographical errors will occur during manual data entry: inadvertent keystrokes can cause digits to be rearranged, dropped, or inserted.  To identify such data entry errors, identifiers often contain a check-digit, which is a redundant checksum consisting of a single digit calculated from the other digits in the identifier. It is typically appended to the end of the identifier, thereby increasing the length of the identifier by one digit.  When an identifier containing a check digit is entered, the computer can verify that the final digit matches the digit predicted by the check digit algorithm. If the two do not match, the identifier can be rejected, resulting in fewer data entry errors. Although there are many check digit methods, a common algorithm is the Luhn algorithm, which is well described [6,7].

## 7.1 Design and implementation of identifiers

Identifiers are needed for numerous components of the health system, including individual persons, specific facilities, individual patient visits, individual records or documents, specific lab results and the like. A national identifier has both merit and risk regardless of whether paper or electronic technologies are used for data collection. Because it is difficult to alter national systems once they are established, countries are advised to take care to make the soundest possible decisions at the outset in putting in place a system for individual identifiers for health services.

## 7.1.1 Terminology

In designing a system of individual identifiers for health services, certain terms are especially important:

- *Unique identifier* – An identifier that is unique is one that can never be associated with more than one individual or entity. Thus, if a particular person is assigned an unique identifier, no other person shall have the same number. An unique identifier also avoids "false positives," or the possibility of linking two records that in reality belong to different people. To achieve this, both a sufficiently large number is needed to cover an entire population for the life time of that population, typically 9-12 characters; an issuing authority is also needed, one that can assure issued numbers will not be reused, or that duplicate numbers will never be issued, e.g., the Social Security Administration in the United States. The issuing authority is typically either a single central authority, or a distributed authority, where local offices issue new identifiers in pre-defined blocks of numbers, each limited to the geographic jurisdiction of the local offices.

- *Ubiquitous identifier* – An identifier is ubiquitous if it is appropriately available at every point within the desired scope of service. For example, if a patient is assigned an identifier for all health services, this identifier cannot be said to be ubiquitous if it is available for use in connection with a clinic visit but not available for a subsequent hospitalization. One method of achieving ubiquity is to embed the identifier onto a smart card, or printed bar code, which the patient carries with him/her whenever receiving health services, or by use of biometric identifiers, such as fingerprints. In countries or other areas with sufficient internet service, ubiquity can be achieved via a web-based software application, which permits access to a central database that can return the patient identifier based on either demographic or biometric identification.

- *Universal identifier* – A universal identifier is used for all purposes of personal identification, including but not limited to health care, banking, driver's licenses and the like. Universal identifiers have the benefit of being broadly recognized and often widely accepted, but they also can present a risk of loss of privacy, stigmatization and misuse of data by others, such as authorities or commercial interests.

### 7.1.2 Assignment of unique identifiers

The particular approach to be used for unique identifiers will depend on a variety of factors. The intended scope of use will determine when and by whom an identifier should be assigned to an individual. If an identifier is to be used for all health services, assignment at birth is appropriate. For identifiers that are used solely to monitor antiretroviral service utilization, an individual's enrollment in an antiretroviral treatment program may be the most useful moment to create a personal identifier. The entity empowered to assign identifiers may include government agencies, health clinics, or other entities.

### 7.1.3 Attributes of identifiers

Steps should be taken to ensure that each identifier is unique. Several options may be considered. For example, an unique sequence may be centrally assigned to ensure that no two people have the same identifier. Sequences may also be assigned locally, with an unique local identifier, for individual clinics, municipalities or other assigning entity, prefixed to the individual's assigned sequence. Another option is assignment of a globally unique identifier or GUID, which uses software to assign identifiers that have an infinitesimally small risk of duplication because of their length.

Personal identifiers for health records should be unchanging, simple, acceptable and practical to implement. Ideally, a health services identifier should contain no personal information, like name or birth date, to avoid the risk of breach of confidentiality. Health services identifiers should be inexpensive, transportable and durable.

### 7.1.4 Ensuring ubiquity

To promote continuity of care, individual identifiers for health services should be readily available for appropriate use in all relevant settings. In particular, efforts should be taken to ensure that identifiers capture all record fragments that are, in fact, linked to the individual with a particular identifier.

These identifiers should be addressed as one component of comprehensive national or system-wide standards for health records. Within the context of comprehensive standards, several non-mutually exclusive options are available to promote the ready availability, or ubiquity, of identifiers.

- One option is to have patients carry their personally identifiable information with them at all times, such as on a card on which these records are stored. A protocol for password and authentication should exist in case an identifier is lost. This option can also be implemented via technology such as smart cards.
- Another option is to allow electronic access to personally identifiable information  at all locations, subject to the requirement that the patient grant access through some procedure for

authentication. For example, the patient could be required to carry a health identification card that may be used to permit a local clinic to access personal data. Biometric scanners might also be used to permit clinic access to individual records. Where individual identifiers are a composite of various bits of personal information like name, date of birth or mother's maiden name, the patient could provide clinic staff with sufficient personal information to permit linkage with individual records.

### 7.1.5 Scope of identifier

Individual identifiers for health services may be implemented for a single vertical, disease-specific programme or for the health system as a whole. Limiting data use to particular vertical programmes or settings may minimize the risks associated with a universal identifier, but decisions in this regard should take into account the context in which health services are to be delivered. For a mobile population, for example, ubiquitous access will be preferable to approaches that limit utilization of information to particular settings.

### 7.2 Legal, ethical and cultural considerations

Identification systems involve a balance between the rights of the individual and the interests of the community. Because personal identification systems both reflect and influence the values of the community, it is important to take legal, ethical and cultural context into consideration in designing unique identifiers.

Development of unique identifier systems should take into account applicable legal frameworks, which may potentially affect how information is used, the acceptability of identifiers, or risks associated with health information systems. Most countries, for example, require that health settings report particular conditions to the appropriate governmental authorities. Often, such reporting requirements are for public health purposes, as in the case of mandates for the reporting of communicable diseases or particular diagnoses. Other reporting provisions serve other purposes, such as reporting requirements for drug use or an individual's immigration status. Consideration of the individual and systemic ramifications of unique identifiers will aid in balancing the need to protect individuals against harm with society's interest in the appropriate enforcement of socially beneficial laws.

Health information systems must also balance society's interest in accurate information with the rights of the individual to control the use of his or her personal information. The *UNAIDS/PEPFAR Interim Guidelines* states that individuals should have access to their records and a right to obtain a copy of them. The *Interim Guidelines* also recommend that individuals have the right to retract records from their files and to add corrective comments. Clients should also be allowed to share their records for the purposes of medical research or medical consultation.

Governments are responsible for sound stewardship of personal information. In this regard, governments should take specific steps and implement appropriate systems to preserve and protect records and ensure their ready accessibility for appropriate use. Legal frameworks should specifically limit governmental access to personal information. In no instance should government have the right to obtain access to identifiable information without due process.

### 7.3 National-level identifiers and their relationship to individual identifiers for health services

Under ideal conditions unique identifiers have the potential to significantly enhance patient care and improve medical outcomes. However, in settings where confidentiality protections are weak and comprehensive information standards are not in place, the risks to the individual of having an unique identifier may outweigh the benefits.

Particular attributes of individual countries, as well as the specific characteristics of the identification system under consideration, affect the risk-benefit calculus with respect to unique identifiers. Identifiers that pertain to a single vertical program or specific disease or condition tend to present a greater risk of harm to the individual. An unique identifier for HIV treatment, for example, risks perpetuating and strengthening the stigma associated with HIV infection. Likewise, attaching an unique identifier to particular populations, such as drug users or sex workers, also presents considerable risks of harm to individuals. Program-specific identifiers pose risk because the very presence of the data field or number on a record implies presence of the condition in question. On the other hand, a too broadly available number such as a national identifier poses increase risk of breach of confidentiality because more people will have access to the number.
Certain national characteristics are highly beneficial for unique identifier systems. Countries with democratic traditions and well-institutionalized laws and rights may be more likely to have the legal framework and cultural context that supports a sound health information system. Likewise, the benefits of unique identifiers are more likely to outweigh risks in countries that recognize due process rights for obtaining access to personal data or exhibit a societal consensus regarding medical ethics. Clear governmental commitment to maximize health care access and reduce stigma and discrimination is also consistent with a sound system of unique identifiers.

Countries with an existing national identification system will need to determine whether the pre-existing identifier is appropriate for health services. In jurisdictions where widespread popular concerns exist regarding the security of the national identifier, authorities may wisely opt not to use the pre-existing identifier for health purposes. By contrast, where a national identifier is already working well and is widely accepted by the population, there may be no rational basis for creating an entirely new system for health services only.

In deciding whether to develop an unique identifier for health services, countries should be aware of the opportunity costs of foregoing such a system. Redundancy in labs and other services represents a hidden, albeit potentially important, financial cost in jurisdictions that lack an unique identifier. Another cost is the expenditure of human resources required to perform wasteful or duplicative record-keeping that could be avoided with an unique identifier. The loss of epidemiological intelligence, strategic information, or monitoring and evaluation data associated with the absence of an unique identifier reduces the effectiveness of public health measures, potentially contributing to avoidable morbidity and mortality.

### 7.4 Use of probabilistic matching

Where definitive linkage between a client and his or her health records is not possible, probabilistic matching represents one method that may be used to narrow the pool of candidate matches. This method is appropriate when a patient has lost or cannot provide his or her unique identifier or when no individual identifiers for health services exist or are consistently accurate or available. By narrowing the universe of records that may be linked to the patient, probabilistic matching may

permit human review of these limited records to enable clinic workers to link the client with the appropriate medical records.

Probabilistic matching may also be used to de-duplicate datasets that may include multiple record fragments from a single client. Ensuring that records pertain to a single client is necessary to avoid inaccuracy in health statistics and to ensure that proper care is provided to individual patients. The quality of the data being matched can vary across different settings and countries, and data quality strongly influences the accuracy that any given algorithm can achieve. Data quality generally refers to data characteristics that influence the degree to which a specific data set can produce accurate matching results. Examples of characteristics that influence data quality and therefore matching results include recording errors, missing values, and presence of highly discriminating fields. Although the performance characteristics of matching systems can be characterized in general terms, the specific performance of probabilistic matching strategies must be evaluated against the data contained in the health records of the countries where they are used.

Whether using probabilistic or deterministic matching methods, data-quality assessment typically occurs as an early step before a matching system is initially implemented. Measuring data quality provides valuable information, and an initial analysis of the characteristics of patient matching informs the processes in at least two ways. First, it assesses how effectively each element in a data source can be used for matching and whether the data can fulfill specific, predefined matching-system performance requirements. Second, data-quality assessments can improve matching accuracy by identifying specific data shortcomings to be addressed by data cleanup strategies (data preprocessing).

When evaluating the performance characteristics of a matching system, it is important to understand the tradeoff between false positive and false negative matches. A false positive match occurs when two truly non-matching records are declared to match, while a false negative match occurs when two truly matching records are declared to be a non-match. Matching systems are often tuned to minimize either false positive or false negative matches, depending on which error is deemed to produce a more undesirable event.

For example, health care matching systems are generally tuned to minimize false positive matches, because a false positive match is generally considered to be more undesirable than a false negative. When records for two separate patients are merged generating a false match, it poses a risk for inadvertent disclosure of protected health information and may lead to incorrect treatment decisions. On the other hand, false negatives limit privacy risks, because no information is disclosed, but may also lead to incorrect treatment decisions.

Although eliminating all false positives may be desirable, it is impractical because false positive and false negatives are inversely related. That is, as the false positive rate is reduced, the false negative rate increases. Low false positive rates are paid for by higher false negative rates: to achieve the ideal false positive rate of zero, the false negative rate approaches 100%, resulting in an ineffective matching system. Consequently, it is important that matching-system implementers understand the relationship between false positive and false negative matches for their specific target data. This understanding is necessary to make an informed decision regarding (a) whether their data quality can support the desired performance characteristics, and (b) at which specific choice of false positive and false negative rates to operate their matching system.

## 8.0 Conclusion

Although the global push to expand HIV treatment access in resource-limited settings faces considerable challenges and obstacles, it also presents critical opportunities. One of the most important of these is the chance to establish sound health information systems to maximize care coordination and improve medical outcomes.

While the workshop discussions primarily focused on the potential benefits of individual identifiers for health services with respect to patient care and public health data, development of a sound health information system also represents an important way to engage the communities most affected by the epidemic. By generating clear, timely information on which strategies are most effective in addressing patients' health care needs, unique identifiers can help provide communities with critical strategic information. Where strong, comprehensive health information systems exist, communities are better equipped to engage in dialogue with public policy makers and health researchers to identify the questions that are of particular interest to people most affected by HIV.
Universal access to HIV prevention, treatment, care and support will be achievable and sustainable only if policy-makers are able to maximize the efficiency and effectiveness of HIV expenditures. Unique identifiers can play a pivotal role in providing policy-makers with the information needed to achieve this outcome.

## 9.0 References

1) UNAIDS/PEPFAR. Interim Guidelines on Protecting the Confidentiality and Security of HIV information: Proceedings from a Workshop, 15-17 May 2006, Geneva, Switzerland. 15 May 2007 http://data.unaids.org/pub/manual/2007/confidentiality_security_interim_guideline_15may2007_en.pdf

2) Beck EJ, Mandalia S, Harling G, Santas X, Mosure D, Delay P. Country Guidelines on the Protection of the Confidentiality & Security of HIV Information: do they exist? XVII International AIDS Conference, Mexico City, Mexico 3-8 August 2008, Abstract WEPE1054.

3) Lohse N, Hansen A-B E, Pedersen G, Kronborg G, Gerstoft J, Sørensen H T, Væth M, Obel N. Survival of Persons with and without HIV Infection in Denmark, 1995–2005, Ann Int Med, 2007,146, 87-95

4) Sogaard OS, Lohse N, Gerstoft J, Kronborg G, Ostergaard L, Pedersen C, Pedersen G, Henrik Toft Sørensen HT, Obel N. Hospitalization for Pneumonia among Individuals With and Without HIV Infection, 1995–2007: A Danish Population Based, Nationwide Cohort Study, 2008; 47, 1345-1353

5) Frank L.When an entire country is a cohort. Science 2000; 5462: 2398-2399.

6) Bello, I, Britton, JR, & Kaul, A. (2007). Topics in contemporary mathematics. Belmont, CA: Brooks/Cole

7) Check Digit, Wikipedia: http://en.wikipedia.org/wiki/Check_digit (accessed 9th May 2010)

# Appendix 1 Participants UNAIDS/PEPFAR Health Services Identifier Workshop

## 24–26 February 2009

## List of Participants

**Lama Abou Char**
UNAIDS
Intern, EMP Department
Geneva, Switzerland
aboucharl@unaids.org

**Mark Amexo**
World Health Organization
Health Metrics Network
Geneva, Switzerland
Tel: +41 22 79 14271
amexom@who.int

**Marina Anderson**
Ministry of Health
Division of Information Management
HIV and AIDS Surveillance and Research
Public Health Specialist, Botswana
Tel: + 267 3632216
Mobile: +267 71508963
ndrsng@yahoo.com or
manderson@gov.bw

**Marcelo Araújo de Freitas**
National STD and AIDS Programme/
Ministry of Health, Brazil
Tel: +55 61 34488067
Mobile: +55 61 81239741
Fax: +55 61 34488009
marcelo.freitas@aids.gov.br

**Eddy Beck**
UNAIDS, EMP
Senior Technical Officer
Geneva, Switzerland
Tel: +41 22 791 5521
Fax: 41 22 791 4798
becke@unaids.org

**Sorakij Bhakeecheep**
National Health Security Office
Manager, Fund Management of HIV/AIDS and TB
Thailand
Tel: (+66)-2-831-4025
Mobile: (+66)-84-751-2289
Fax: (+66)-2-831-4004
sorakij.b@nhso.go.th

**Stephane M Bodika**
M & E Surveillance Advisor
BOTUSA,
Centers for Disease Control and Prevention
Gaborone, Botswana
Tel: +267367-2468
Fax: +2 67-3930429
bodikas@bw.cdc.gov

**Philippe Boucher**
World Health Organization
Web Development Officer EIP, KMS
Geneva, Switzerland
Tel: +41 22 791 3688
boucherp@who.int

**Tetyana Boyko**
Ukrainian AIDS Centre
Head of the Department of Organization and
Methodology
Kyiv, Ukraine
Tel: +38-050-227-9442
Tetyana@list.ru

**Evan Collins**
Ontario HIV Treatment Network (OHTN)
President, Board of Directors
Ontario, Canada
Tel: 1-416-603-6027
Fax: 1-416-922-2018
ecollins@interlog.com

**Agnes Cordero**
UNAIDS, EMP
Administrative Assistant
Geneva, Switzerland
Tel :+41 22 791 4555
corderoa@unaids.org

**Beatriz De Faria Leao**
Zilics Health Information Systems
MD, PhD - Health Standards Architect
Brazil
Tel: + 55 -11 -81141617
Fax: +55-11-30551600
beatriz.leao@zilics.com.br

**Maria Lorena Di Giano**
Argentinean Network of Women Living with HIV
Lawyer- Human Rights and HIV/AIDS Advisor
Mar del Plata, Argentina
Tel: +54 223 4711822
loredigiano@hotmail.com or
lorenadigiano@gmail.com

**Shaun J. Grannis**
The Regenstrief Institute
Medical Informatics Research Scientist
Indiana University School of Medicine
Assistant Professor of Family Medicine
United States
Tel: (317) 423-5523
Fax: (317) 423-5695
sgrannis@regenstrief.org

**Catherine Hankins**
UNAIDS, EMP
Chief Scientific Officer
Geneva, Switzerland
Tel: +41227913865
Fax: +41227914746
hankinsc@unaids.org

**Mike Isbell**
Rapporteur
New York, United States
Tel: + 1 212 924 3935
mtisbell@aol.com

**Neil Jacobs**
Centers for Disease Control
HMIS Specialist, South Africa
Tel: +27-12-424-9000
Fax: +27-12-346-0935 9045# (e-Fax)
jacobsn@sa.cdc.gov

**Andreas Jahn**
Department for HIV & AIDS, Ministry of Health
Malawi, Central Monitoring & Evaluation Division
Tel: +265 9 387 418
Fax: +265 1 789 365
drandreasjahn@gmail.com

**Sergon Kibet**
Ministry of Public Health and Sanitation
Head, Health Management Information system
Kenya
Tel: +254 722 659 568, +254 203 532 622
kibetsergon@yahoo.com

**Svilen Konov**
HIV I-Base
United Kingdom
Tel: + 020 7407 8488
Facsimile: +020 7407 8489
svilen.konov@i-base.org.uk

**Mark Landry**
Office of the U.S. Global AIDS Coordinator
U.S. President's Emergency Plan for AIDS Relief
Senior Informatics Specialist
Washington, DC, United States of America
Tel: 202.663.2355
Fax: 202.663.2943
LandryMS@state.gov

**Ruth Macklin**
Department of Epidemiology and Social Medicine
Albert Einstein College of Medicine 1300 Morris
Park Avenue
New York, United States of America
Telephone: 1 718 430 3574
Facsimile: 1 718 430 8780
Macklin@aecom.yu.edu

**Faye Malitz**
U.S. DHHS; Health Resources and Services
Administration; HIV/AIDS Bureau
Director, Division of Science and Policy
United States
Tel: (301) 443-3259
Fax: (301) 594-2511
Fmalitz@hrsa.gov

**Dorina Maris**
USAID
Technical Advisor, Health Management Information
Systems
United States
Tel: 202-712-5354
Fax: 202-216-3312
dmaris@usaid.gov

**Debra Mosure**
Global AIDS Program
Centers for Disease Control and Prevention
Deputy Associate Director for Science
Atlanta, United States
Tel: 404-639-1857
Fax: 404-639-4268
DMosure@cdc.gov

**Lydia Mungherera**
The AIDS Support Organization (TASO)
TASO Headquarters Mulago
Program Officer/ Training
Kampala, Uganda
Tel: +256 41 532580
Fax: +256 41 541288
Mobile: +256 77 448102
munghereral@tasouganda.org ;
lydiamng@yahoo.co.uk

**Niels Obel**
Danish HIV Cohort Study, Copenhagen University
Hospital
Consultant, MD, PhD
Copenhagen, Denmark
Tel: (+45) 35457730
Fax: (+45) 35457750
niels.obel@rh.regionh.dk

**Rachel Ong**
Asia Pacific Network of People Living with
HIV/AIDS (APN+)
Advisor
Thailand
Tel: +662-2557477
Fax: +662-2557479
rachel.ong.pcb@gmail.com

**Heather Patrick**
Global AIDS Program
Centers for Disease Control and Prevention
Data Support Team
Atlanta, United States
htx7@cdc.gov

**Claudia Risso de Araujo Lima**
DATASUS / Health Ministry
Brazil
Tel: (55 21) 3985-7216
Fax: (55 21) 3985-7240
crisso@datasus.gov.br

**Xen Santas**
Global AIDS Program
Centers for Disease Control and Prevention
Lead, Data Support Team
Atlanta, United States of America
Tel: + 1 404 639 2036
xms1@cdc.gov

**Mark Shields**
Centers for Disease Control and Prevention
Branch Chief, Epidemiology and Strategic
Information
Zambia
Mobile: +260 977 790786
E-fax: +202-318-4062
mark@healthyafrica.net

**Noami Shiyango**
Kenyatta National Hospital
National AIDS/STD Control Program (NASCOP)
Kenya
Tel: +254722562826
naomishiyonga@yahoo.com.

**Sisay Sirgu**
World Health Organization
IMAI, HIV Department
Technical officer
Geneva, Switzerland
Tel: +41227913536
Fax: +41227911575
sirgus@who.int

**Patricia Sweeney Hardy**
Division of HIV/AIDS Prevention
Centers for Disease Control and Prevention
Epidemiologist, United States
Tel: 404-639-2047
Fax: 404-639-2980
pas3@cdc.gov

**Jean Michel Tassie**
World Health Organization
HTM/HIV/SIR, Medical Officer
Geneva, Switzerland
Tel: +41 22 79 13470
tassiej@who.int

**Achara Teeraratkul**
Chief, Strategic Information, Global AIDS
Program/Thailand
CDC Southeast Asia Regional Office
Ministry of Public Health
Tiwanon Rd.
Nonthaburi 11000, Thailand
Tel: +662 580 0669 ext. 531
agt4@th.cdc.gov

**Derrick Tlhoiwe**
ARV Programme Manager
MOH Botswana
Tel: + 267 3632216
drderricktlhoiwe@yahoo.com

**Steven Uggowitzer**
World Health Organization
Technical Officer Health Metrics Network
Geneva, Switzerland
Tel: +41 22 79 11678
uggowitzers@who.int

**Olga Varetska**
International HIV/AIDS Alliance in Ukraine
Head of Team, Monitoring and Evaluation
Kyiv, Ukraine
Tel.: (+380 44) 490-5485 Ext. 130
varetska@aidsalliance.org.ua

**Anandi Yuvaraj**
International Community of Women Living with
HIV/AIDS
Asia Pacific Regional Coordinator
Thailand
Tel: +66 22 55 74 77-78
Fax: +66 22 55 74 79
anandiy@hotmail.com

# Appendix 2 Annotated Bibliography

## 5.0 FURTHER READING – LEGAL, ETHICAL, PRIVACY CONSIDERATIONS...............62

## 6.0 FURTHER READING – IMPLEMENTATION/CASE STUDIES .......................................65

# INTRODUCTION

## *Purpose of the Annotated Bibliography*

An effective response to the HIV epidemic is longitudinal in nature and multisectoral in scope. Optimal service provision requires health service providers to collect, store and manage information on the same individual over time and across different points of service delivery. Sensitive information on HIV-positive patients is being collected in the absence of policies and procedures that protect the identity of people using health and other services. Recognizing that stigma and discrimination continue to be important drivers of the HIV epidemic, members from the Joint United Nations Programme on HIV/AIDS (UNAIDS) and the President's Emergency Plan for AIDS Relief (PEPFAR) convened a workshop in 2006 that produced the Interim Guidelines on Protecting the Confidentiality and Security of HIV Information.[2]

Collecting accurate patient data is complicated by the multitude of data sources, the disparity in technology, inconsistencies in data storage formats between facilities and over time, shifting data needs, and the absence of unique patient identifiers. Building on the 2006 Confidentiality and Security of HIV Information Workshop, UNAIDS and PEPFAR are now convening the Health Services Identifier Workshop in Montreux, Switzerland, 24–26 February 2009. The goal of this workshop is to bring together relevant experts to address these issues and to produce guidelines for countries on the use of health service identifiers to accurately identify patients and uniquely match patient records for providing, monitoring, and evaluating HIV services.

This annotated bibliography was developed to provide a review of existing material about health services identifiers and present different countries' experiences with such identifiers. The annotated bibliography is intended to be used as a resource for the workshop, presenting material that workshop participants can review and critique and use to inform the development of guidelines for the design and implementation of health service identifiers in middle- and lower-income countries.

## *Approach to the Annotated Bibliography*

Macro International Inc. was contracted to do the literature search and put together this annotated bibliography. The literature search was guided by health management information system specialists from the Centers for Disease Control and Prevention and other partners, and included a systematic assessment of both published literature and industry information found through Web sites. The literature search was conducted using key databases, particularly MEDLINE and PubMed, online search engines such as Google, and citations found in articles.

The initial search terms were derived from the draft agenda for the Health Services Identifier Workshop and included *unique health identifiers, HIV longitudinal monitoring, privacy, disease registries, HIV disease registries, biometric readers, smartcards, public key infrastructure,* and *probabilistic matching*, in various combinations. Articles located through these methods provided additional search terms.

---

[2] UNAIDS/PEPFAR. (2007, May 15). Interim guidelines on protecting the confidentiality and security of HIV information. *Proceedings from a Workshop*, 15–17 May 2006, Geneva, Switzerland.
http://data.unaids.org/pub/manual/2007/confidentiality_security_interim_guideline_15may2007_en.pdf

Located articles were distributed to each member of the review team. Each article or document was reviewed, selected for inclusion/exclusion, and classified independently and then reviewed and discussed with all three team members. The selection of classification terms was an emergent process. Review of the literature highlighted three distinct methods for uniquely identifying patients: using unique patient identifiers, using technology such as smartcards or biometrics, and matching patient records from different databases using a combination of non-unique identifiers. These three approaches are used as broad categories to classify the literature. In addition, articles about privacy, security, legality, and ethical issues around health identifiers are classified separately. Examples of implementation in different countries, or implementation guidance documents, are included as a separate category.

## *Annotated Bibliography Content and Format*

The annotated bibliography is divided into two sections—recommended reading and further reading. The list of 15 recommended readings includes articles from all five categories (unique identifiers, record linkage, technology, privacy, and implementation). These articles are recommended because they include concise overviews of the key topics with sufficient level of detail. The articles included as further reading provide additional context and detail, as well as differing viewpoints.

The bibliographic reference for each entry appears in American Psychological Association (APA) format. Each entry, listed in alphabetical order, includes the title and full reference of the article, an abstract, and a URL to locate the article online. If an article was missing an abstract, the reviewer has summarized the article and included a brief synopsis in lieu of an abstract.

## *Access to the articles*

The annotated bibliography shows a URL address for each document listed. In most cases the hyperlink allows direct access to the article. However, some material is available only by subscription. In these cases the hyperlink is only to an abstract.

# 1.0 GENERAL RECOMMENDED READING

1.1.   ASTM International (2007a). *ASTM E-1714-00 standard guide for properties of a universal healthcare identifier (UHID)*. West Conshohocken, PA: Author.

This guide covers a set of requirements outlining the properties required to create a universal healthcare identifier (UHID) system. The document sets forth the fundamental considerations for a UHID that can support at least four basic functions: (a) Positive identification of patients when clinical care is rendered, (b) Automated linkage of various computer-based records on the same patient for the creation of lifelong electronic health care files, (c) Provision of a mechanism to support data security for the protection of privileged clinical information, and (d) The use of technology for patient records handling to keep health care operating costs at a minimum. This standard does not purport to address all safety concerns, if any, associated with its use.

http://www.astm.org/Standards/E1714.htm


1.2.   ASTM International. (2007b). *ASTM E2553 guide for implementation of a voluntary universal healthcare identification system.* West Conshohocken, PA: Author.

This document describes the implementation principles needed to create a Voluntary Universal Healthcare Identification (VUHID) system. The purpose of this system is to enable unambiguous identification of individuals in order to facilitate the delivery of healthcare. The VUHID system should be dedicated exclusively to the needs and functions of healthcare. The VUHID system is designed to represent no, or at least minimal, increased risk to healthcare privacy and security. The system should be as cost effective as possible. The system must be created and maintained in a way to provide sustained benefit to healthcare. The system should be designed and implemented in a manner that ensures that it can operate indefinitely. This standard does not purport to address all of the safety concerns, if any, associated with its use.

http://www.astm.org/Standards/E2553.htm


1.3.   Churches, T., & Christen, P. (2004). Some methods for blindfolded record linkage. BMC Medical Informatics and Decision Making, 4, 9.

Background: The linkage of records which refer to the same entity in separate data collections is a common requirement in public health and biomedical research. Traditionally, record linkage techniques have required that all the identifying data in which links are sought be revealed to at least one party, often a third party. This necessarily invades personal privacy and requires complete trust in the intentions of that party and their ability to maintain security and confidentiality.

Methods: A method is described which permits the calculation of a general similarity measure, the n-gram score, without having to reveal the data being compared, albeit at some cost in computation and data communication. This method can be combined with public key cryptography and automatic estimation of linkage model parameters to create an overall system for blindfolded record linkage.

Results: The system described offers good protection against misdeeds or security failures by any one party, but remains vulnerable to collusion between or simultaneous compromise of two or more parties involved in the linkage operation. In order to reduce the likelihood of this, the use of last-minute

allocation of tasks to substitutable servers is proposed.

Conclusion: Although the protocols described in this paper are not unconditionally secure, they do suggest the feasibility, with the aid of modern cryptographic techniques and high-speed communication networks, of a general purpose probabilistic record linkage system which permits record linkage studies to be carried out with negligible risk of invasion of personal privacy.

http://www.biomedcentral.com/1472-6947/4/9


1.4.    Hieb, B. R. (2006). The case for a voluntary national healthcare identifier. *Journal of ASTM International, 3*(2).

This paper describes a proposed system to create a voluntary national healthcare identification (VNHID) system for the United States. This system would be implemented in addition to the national linkage mechanisms currently being proposed as part of a national health information network. It will provide demonstrable improvements in the privacy, security, and efficiency of the system. It would also eliminate a significant set of errors inherent in the currently proposed health information linkage system. The proposed VNHID system is able to meet the vast majority of the objections that have previously been raised concerning a national healthcare identifier on the basis of privacy concerns. The system has the potential to be implemented rapidly and would not require the development of a national consensus prior to implementation.

http://www.astm.org/JOURNALS/JAI/PAGES/JAI13891.htm


1.5.    Hillestad, R., Bigelow, J. H., Chaudhry, B., Dreyer, P., Greenberg, M. D., Meili, R. C., et. al. (2008). *Identity crisis: An examination of the costs and benefits of an unique patient identifier for the U.S. health care system* (RAND MG-753-HLTH). Santa Monica, CA: Rand Corporation.

Correctly linking patients to their health data is a vital step in quality health care. The two primary approaches to this linking are the unique patient identifier (UPI) and statistical matching based on multiple personal attributes, such as name, address, and Social Security number (SSN). Lacking a UPI, most of the U.S. health care system uses statistical matching methods. There are important health, efficiency, security, and safety reasons for moving the country away from the inherent uncertainties of statistical approaches and toward a UPI for health care. In this monograph, we compare the linking alternatives on the basis of errors, cost, privacy and information security, and political considerations. We also discuss operational efficiency, ease of implementation, and some implications for improved health care.

http://www.rand.org/pubs/monographs/2008/RAND_MG753.pdf


1.6.    McMahon, D. J. (2008). The future of privacy in a unified national health information infrastructure. *Seton Hall Law Review, 38*, 787.

This comment begins with an overview of the current state of healthcare privacy law and the need for adequate privacy protection. Part III then describes and analyzes selected bills which are paradigmatic of the various approaches that Congress currently contemplates. Part IV examines different methods of privacy protection available to supplement these bills. Part IV also argues that the most effective way to protect personal privacy in a national health information infrastructure is through a multilayered approach which utilizes a new property right in personal information along with contractual and tort-

based protection.

http://www.ncbi.nlm.nih.gov/pubmed/18623904

1.7.   Meray, N., Reitsma, J. B., Ravelli, A. C., & Bonsel, G. J. (2007). Probabilistic record linkage is a valid and transparent tool to combine databases without a patient identification number. *Journal of Clinical Epidemiology, 60*(9), 883–891.

Objective: To describe the technical approach and subsequent validation of the probabilistic linkage of the three anonymous, population-based Dutch Perinatal Registries (LVR1 of midwives, LVR2 of obstetricians, and LNR of pediatricians/neonatologists). These registries do not share an unique identification number.

Study Design and Setting: A combination of probabilistic and deterministic record linkage techniques was applied using information about the mother, delivery, and child(ren) to link three known registries. Rewards for agreement and penalties for disagreement between corresponding variables were calculated based on the observed patterns of agreement and disagreement using maximum likelihood estimation. Special measures were developed to overcome linking difficulties in twins. A subsample of linked and nonlinked pairs was validated.

Results: Independent validation confirmed that the procedure successfully linked the three Dutch perinatal registries despite nontrivial error rates in the linking variables.

Conclusions: Probabilistic linkage techniques allowed the creation of a high-quality linked database from crude registry data. The developed procedures are generally applicable in linkage of health data with partially identifying information. They provide useful source date even if cohorts are only partly overlapping and if, within the cohort, multiple entities and twins exist.

http://www.jclinepi.com/article/S0895-4356(06)00500-2/pdf

1.8.   Meux, E. (1994). Encrypting personal identifiers. *Health Services Research, 29*(2), 247–256.

Study Setting: A statewide patient discharge database contained only one unique identifier: the social security number (SSN). A method was developed to transform (encrypt) the SSN so that it could be made publicly available, for purposes of linking discharge records, without revealing the SSN itself. The method of encrypting the SSN into a Record Linkage Number (RLN) is described.

Principal Findings: The same RLN will always result from the same SSN; it is highly improbable that the same RLN would be produced by two different SSNs; the SSN cannot be derived from the RLN, even given access to the encryption program; the encryption method cannot be determined through knowledge of a number of SSN/RLN combinations; and the method can be described, evaluated, and adapted for use by other researchers without compromising confidentiality of the RLNs resulting from the method.

http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=1070001&blobtype=pdf

1.9.   National E-Health Transition Authority (2006). *Privacy blueprint–Unique healthcare identifiers: Individual healthcare identifier and healthcare provider identifier*. Sydney, Australia: Author.

The Privacy Blueprint on NEHTA's Unique Healthcare Identifiers (UHI) program establishes a clear framework to consider privacy issues raised by the development and implementation of national

healthcare identification infrastructure. It summarizes NEHTA's progress to date in managing privacy issues arising from the UHI Service and sets out an action plan for future work.

The UHI Service consists of two discrete identifiers—a Healthcare Provider Identifier for healthcare providers and healthcare organizations, and an Individual Healthcare Identifier for individual consumers. This Privacy Blueprint outlines the nature and function of the UHI Service, and identifies key participants in the proposed system as well as the personal and health information involved and how it will be used.

Importantly, this Privacy Blueprint is largely based on a generic Australian privacy analysis, reflecting the fact that it was not known whether the UHI Organization responsible for managing the UHI Service would be a public or a private sector organization. This is critical, as the nature of the organization determines whether it is subject to the public or private sector provisions of the Privacy Act 1988. While the public and private sector privacy principles are similar, there are key differences in coverage and the way they are structured that will impact the privacy analysis undertaken.

http://www.nehta.gov.au/index.php?option=com_docman&task=doc_view&gid=148&Itemid=139

1.10.   Netter, W. (2003). Curing the unique health identifier: Reconciliation of new technology and privacy rights. *Jurimetrics, 43*(2), 165–186.

The Health Insurance and Portability and Accountability Act has mandated the assignment of a universal individual health identifier in 2003. Such an identifier can increase patient confidentiality, improve patient care, lower the cost of services to patients, enhance administrative efficiency, and increase the opportunity for medical research. Nevertheless, national identification systems raise concerns about confidentiality and privacy. Instead of a mandatory, government-assigned number, this article proposes a technological multitiered system that would be administered by a mixed government and private entity. Consumers could voluntarily opt- in to the system.

http://medscapecrm.org/medline/abstract/15156882

1.11.   Office of the National Coordinator for Health Information Technology (2008). *Nationwide privacy and security framework for electronic exchange of individually identifiable health information*. Rockville, MD: U.S. Department of Health and Human Services.

The principles presented in this document establish a single, consistent approach to address the privacy and security challenges related to electronic health information exchange through a network for all persons, regardless of the legal framework that may apply to a particular organization. The goal of this effort is to establish a policy framework for electronic health information exchange that can help guide the Nation's adoption of health information technologies and help improve the availability of health information and health care quality. The principles have been designed to establish the roles of individuals and the responsibilities of those who possess and exchange electronically identifiable health information through a network.

http://www.hhs.gov/healthit/documents/NationwidePS_Framework.pdf

1.12.   Public Health Informatics Institute (2006). *The unique records portfolio: A guide to resolving duplicate records in health information systems*. Decatur, GA: Author.

The portfolio explores the problem of unique person record identification within the integration of individual/disparate databases (the source databases). Linking data from disparate information systems forces an organization to be explicit about the intended uses of the linked data, to understand the risks associated with inaccurately matching data, and to establish a strategy that supports the goals of the record matching measured against its inherent complications and risks. The strategies developed to address duplicate records are often referred to as deduplication strategies. The concepts, examples, and tools in the Unique Records Portfolio are intended to help public health agencies and health care organizations address the deduplication challenge. The portfolio also articulates the challenges and solutions to developing deduplication strategies, and describes the implications that various approaches have on data use. Case examples of the deduplication strategies of several state integrated information systems provide real-world experience, and hands-on tools assist managers in thinking through the various aspects of the decisions they need to make.

http://www.phii.org/resources/UniqueRecordsPortfolio.asp

1.13.   Sauleau, E. A., Paumier, J. P., & Buemi, A. (2005). Medical record linkage in health information systems by approximate string matching and clustering. *BMC Medical Informatics and Decision Making, 5*, 32.

Background: Multiplication of data sources within heterogeneous healthcare information systems always results in redundant information, split among multiple databases. Our objective is to detect exact and approximate duplicates within identity records, in order to attain a better quality of information and to permit cross-linkage among stand-alone and clustered databases. Furthermore, we need to assist human decision making by computing a value reflecting identity proximity.

Methods: The proposed method is in three steps. The first step is to standardise and index elementary identity fields, using blocking variables, in order to speed up information analysis. The second is to match similar pair records, relying on a global similarity value taken from the Porter-Jaro-Winkler algorithm. The third is to create clusters of coherent related records, using graph drawing, agglomerative clustering methods, and partitioning methods.

Results: The batch analysis of 300,000 "supposedly" distinct identities isolates 240,000 true unique records, 24,000 duplicates (clusters composed of 2 records), and 3,000 clusters whose size is greater than or equal to 3 records.

Conclusion: Duplicate-free databases, used in conjunction with relevant indexes and similarity values, allow immediate (i.e., real-time) proximity detection when inserting a new identity.

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1274322

1.14.   Stewart, S. P., Arellano, M. B., & Simborg, D. W. (1984). Optimal patient identification system. *Journal of the American Medical Record Association, 55*(8), 23–27.

Health facilities continually seek ways to provide services more efficiently. Linking all facets of patient care through a well-designed patient identification (PID) system can be expected to increase the efficiency of patient care. Linking patient service data to financial data allows administrators to analyze costs. Efficiency will be improved if all hospital departments adopt a common patient identification number as the means of patient identification. This article offers guidelines for implementing a cost-effective computerized patient identification system.

http://www.ncbi.nlm.nih.gov/pubmed/10310711


1.15.    Tierney, W. M., Beck, E. J., Gardner, R. M., Musick, B., Shields, M., Shiyonga, N. M., et al. (2006). Viewpoint: A pragmatic approach to constructing a minimum data set for care of patients with HIV in developing countries. Journal of the American Medical Informatics Association, 13(3), 253–260.

Providing quality health care requires access to continuous patient data that developing countries often lack. A panel of medical informatics specialists, clinical human immunodeficiency virus (HIV) specialists, and program managers suggests a minimum data set for supporting the management and monitoring of patients with HIV and their care programs in developing countries. The proposed minimum data set consists of data for registration and scheduling, monitoring and improving practice management, and describing clinical encounters and clinical care. Data should be numeric or coded using standard definitions and minimal free text. To enhance accuracy, efficiency, and availability, data should be recorded electronically by those generating them. Data elements must be sufficiently detailed to support clinical algorithms/guidelines and aggregation into broader categories for consumption by higher level users (e.g., national and international health care agencies). The proposed minimum data set will evolve over time as funding increases, care protocols change, and additional tests and treatments become available for HIV-infected patients in developing countries.

http://www.jamia.org/cgi/content/abstract/13/3/253

# 2.0 FURTHER READING – UNIQUE IDENTIFIERS

2.1     Appavu, S. I. (1997). Analysis of unique patient identifier options final report. Prepared for the U.S. Department of Health and Human Services.

The Health Insurance Portability and Accountability Act of 1996 requires the Secretary of the U.S. Department of Health and Human Services (HHS) to adopt standards for Unique Health Identifiers to identify individuals in addition to providers, health plans, and employers. The industry has put forth several options for the Unique Patient Identifier. The objective of this study is to perform an analysis of the various Unique Patient Identifier options that are available for use in health care. The result of this analysis will facilitate and support the recommendation to be made to the Secretary of HHS by the National Committee on Vital and Health Statistics.

http://ncvhs.hhs.gov/app0.htm


2.2     Appavu, S. I. (1999). Unique patient identifiers: What are the options? Journal of AHIMA, 70(9), 50–57.
In 1997, the U.S. Department of Health and Human Services commissioned a study to analyze the various patient identification systems available. The study consisted of an objective analysis of the various unique patient identifier (UPI) options available for use in the healthcare system using four levels of criteria: conceptual, operational, component, and functional. The study also examined industry requirements and looked at the needs of the industry as a whole. This article presents a brief overview of the analysis described in the study.

http://www.ncbi.nlm.nih.gov/pubmed/10977406


2.3     Australia Standard Committee IT-014 Health Informatics. (2006). Health care client identification (AS 5017-2006). Sydney: Standards Australia.

The objective of this standard is to provide the health industry with a specific standard for healthcare client identification for clinical and administrative data management purposes (data structure and specification) which promotes uniformly good practice in identifying individuals and recording identifying data so as to ensure that each individual's health record will be associated with that individual and no other. The standard also provides the basis for future linkage of data as authorised by law and appropriate for clinical management of patients and statistical research purposes.

http://www.saiglobal.com/PDFTemp/Previews/OSH/as/as5000/5000/5017-2006.pdf

2.4     Fernandes, L., & O'Connor, M. (2008). Patient identification in three acts. *Journal of AHIMA, 79*(4), 46–49

This article discusses three solutions for patient identification which appear to be emerging ahead of the pack—probabilistic matching, national identifier, and voluntary Universal Health Identifier (VUHID). The article notes recent developments, such as the U.S. presidential election and the release of the RAND report, which may have an impact on patient identification standards.

 http://library.ahima.org/xpedio/groups/public/documents/ahima/bok1_037463.hcsp

2.5    France, F. H., De Clercq, E., & Bangels, M. (2005). Purposes of health identification cards in Belgium. In R. Engelbrecht, A. Geissbuhler, & C. Lovis (Eds.), Connecting medical informatics and bio-informatics. Amsterdam: IOS Press.

Although other alternatives might exist, identification cards have been chosen as an acceptable and adequate tool to be used to identify patients and health professionals. They are planned for a digital signature and for access to electronic health records as well as for health information exchange and for database querying. Local applications might exist independently, but the Federal State has now developed Be- Health, a platform for health professionals, social security personnel, as well as the great public to facilitate a common access to some health data. Security conditions have been defined and are described.
http://iospress.metapress.com/content/0v63l16jdx50py30

2.6    Gabrieli, E. R. (1993). Guide for unique healthcare identifier model. Journal of Clinical Computing, 21(5), 101–139.

This guide presents the current understanding of the role and construction of an unique Healthcare Identifier (UHI) intended to be strictly limited to the use within the American healthcare system. The guide (a) sets forth the fundamental criteria for a UHI that can facilitate linking computer-based records on the same patient, (b) provides for data security measures for confidential clinical data, and (c) identifies patients receiving clinical care.
http://www.ncbi.nlm.nih.gov/pubmed/10126775

2.7    McBride, A. (1998). The identification standard explained. Ottawa, Ontario: Canadian Institute for Health Information.

The Health Identification Standard for Individuals in Canada was approved by the Canadian Institute for Health Information Partnership for Health Informatics/Telematics as a trial use standard in 1998. This paper discusses some of the many aspects of identifying individuals, and describes the legislative basis for the standard. Also discussed are the topics of documents supporting identification and legal names. An explanation of the objectives of the standard is provided along with example scenarios of how the standard would work in practice. In addition, the status of two projects involving the implementation of very similar identification standards is reviewed. The paper concludes with discussions on the relationship between documentary and biometric identification and why the standard should be adopted and implemented.

http://secure.cihi.ca/cihiweb/en/downloads/infostand_unique_e_wg2idstd.pdf

2.8    Quantin, C., Allaert, F. A., Avillach, P., Fassa, M., Riandey, B., Trouessin, G., et al. (2008). Building application-related patient identifiers: What solution for a European country? International Journal of Telemedicine and Applications.

We propose a method utilizing a derived social security number with the same reliability as the social security number. We show the anonymity techniques classically based on unidirectional hash functions (such as the secure hash algorithm (SHA-2) function that can guarantee the security, quality, and reliability of information if these techniques are applied to the Social Security Number). Hashing produces a strictly anonymous code that is always the same for a given individual, and thus enables patient data to be linked. Different solutions are developed and proposed in this article. Hashing the social security number will make it possible to link the information in the personal medical file to other

national health information sources with the aim of completing or validating the personal medical record or conducting epidemiological and clinical research. This data linkage would meet the anonymous data requirements of the European directive on data protection.

http://www.hindawi.com/GetArticle.aspx?doi=10.1155/2008/678302


2.9    Rollins, G. (2007). This year's models: A look at patient ID in the four newly demonstrated NHIN prototypes. Journal of AHIMA, 78(3), 34–37.

This winter the healthcare industry got a first look at four prototypes for a nationwide healthcare data exchange network. Created under contract to the federal government, the demonstration projects paired technology developers and healthcare providers in 12 communities across the United States. As the contractors wound up their first year of work and prepared to unveil their models, the Journal of AHIMA spoke with them about their individual approaches to arguably the biggest initial hurdles in widespread data exchange–patient identification and record linkage. This article describes the prototypes and their unique approaches.

http://library.ahima.org/xpedio/groups/public/documents/ahima/bok1_033608.hcsp

# 3.0 FURTHER READING – RECORD LINKAGE

3.1    Alemi, F., Loaiza, F., & Vang, J. (2007). Probabilistic master lists: Integration of patient records from different databases when unique patient identifier is missing. Health Care Management Science, 10, 95–104.

We show how Bayesian probability models can be used to integrate two databases, one of which does not have a key for uniquely identifying clients (e.g., social security number or medical record number). The analyst selects a set of imperfect identifiers (last visit diagnosis, first name, etc.). The algorithm assesses the likelihood ratio associated with the identifier from the database of known cases. It estimates the probability that two records belong to the same client from the likelihood ratios.

We test that the procedure is effective by examining data from the Medical Expenditure Panel Survey (MEPS) Population Characteristics data set, a publicly available data set. We randomly selected 1,000 cases for training data set—these constituted the known cases. The algorithm was used to identify if 100 cases not in the training data set would be misclassified in terms of being a case in the training set or a new case. With 12 fields as identifiers, all 100 cases were correctly classified as new cases. We also selected 100 known cases from the training set and asked the algorithm to classify these cases. Again, all 100 cases were correctly classified.

These data suggest the accuracy of our automated and mathematical procedure to merge data from two different data sets without the presence of an unique identifier. The algorithm uses imperfect and overlapping clues to reidentify cases from information not typically considered to be a patient identifier.

http://www.springerlink.com/content/c0n30483l32421r6


3.2    Brum, L., & Kupek, E. (2005). Record linkage and capture-recapture estimates for underreporting of human leptospirosis in a Brazilian health district. The Brazilian Journal of Infectious Diseases, 9(6), 515–520.

Record linkage and capture-recapture models were used to estimate the number of cases of human leptospirosis in the health district of Santa Maria, RS in southern Brazil. Twelve months of laboratory, hospital, and epidemiological surveillance data were matched by name, age, residence, and the month of diagnosis. Only laboratory-confirmed cases were considered. The record linkage revealed more than 20 times more cases than the official estimate for the health district, indicating a leptospirosis epidemic, with an annual incidence of more than 3 per 1,000 inhabitants and a case fatality of 0.37%. Severe cases were predominantly found through hospital records, overlapping to some extent with the epidemiological surveillance data, whereas less severe cases were found almost exclusively through laboratory logs. Different combinations of data sources influenced the detection rate for low versus high severity cases. Based on log-linear capture-recapture models, stratified by case severity and taking into account possible dependencies between the data sources, an insignificant number of cases were missed by all sources.

http://www.scielo.br/pdf/bjid/v9n6/a11v09n6.pdf


3.3    Christen, P. (2008). Febrl: A freely available record linkage system with a graphical user interface. Proceedings of the Second Australasian Workshop on Health Data and Knowledge Management, Vol. 80. Sydney, NSW, Australia: Australian Computer Society.

Record or data linkage is an important enabling technology in the health sector, as linked data is a cost-effective resource that can help to improve research into health policies, detect adverse drug reactions, reduce costs, and uncover fraud within the health system. Significant advances, mostly originating from data mining and machine learning, have been made in recent years in many areas of record linkage techniques. Most of these new methods are not yet implemented in current record linkage systems, or are hidden within 'black box' commercial software. This makes it difficult for users to learn about new record linkage techniques, as well as to compare existing linkage techniques with new ones. What is required are flexible tools that enable users to experiment with new record linkage techniques at low costs.

This paper describes the Febrl (Freely Extensible Biomedical Record Linkage) system, which is available under an open source software licence. It contains many recently developed advanced techniques for data cleaning and standardisation, indexing (blocking), field comparison, and record pair classification, and encapsulates them into a graphical user interface. Febrl can be seen as a training tool suitable for users to learn and experiment with both traditional and new record linkage techniques, as well as for practitioners to conduct linkages with data sets containing up to several hundred thousand records.

http://crpit.com/confpapers/CRPITV80Christen.pdf


3.4    Colias, M. (2005). Disease registries. Hospitals & Health Networks, 79(2), 62–68.

The article provides a brief overview of disease registries, describing their three main uses, the benefits and challenges of implementing a disease registry, and presents anecdotal evidence that suggests registries vastly improve providers' ability to manage chronic conditions.

http://www.hhnmag.com/hhnmag/html/2005_Index.html


3.5    Dunn, H. L. (1946). Record linkage. American Journal of Public Health, 36, 1412–1416.

In this article, the author introduces the term "record linkage" to express the concept of collating health care records into a cumulative personal file, starting at birth and ending at death. The article emphasizes the value of linked files at different levels—for the individual, for registrars of vital records, and for health, welfare, and other types of organizations.

http://www.ajph.org/cgi/reprint/36/12/1412


3.6    Fellegi, I. P., & Sunter, S. B. (1969). A theory of record linkage. Journal of the American Statistical Association, 64(328), 1183–1210.

A mathematical model is developed to provide a theoretical framework for a computer-oriented solution to the problem of recognizing those records in two files which represent identical persons, objects or events (said to be matched).

A comparison is to be made between the recorded characteristics and values in two records (one from each file) and a decision made as to whether or not the members of the comparison-pair represent the same person or event, or whether there is insufficient evidence to justify either of these decisions as stipulated levels of error.

A theorem describing the construction and properties of the optimal linkage rule and two corollaries to

the theorem which make it a practical working tool are given.

http://www.jstor.org/pss/2286061

3.7     Gill, L. (2001). Methods for automatic record matching and linking and their use in national statistics. Newport, UK: Office for National Statistics.

This report concentrates on the use of record linkage for statistical purposes only, to produce summaries and statistics. Many of the administrative and survey data sets are collected under legislative framework and are subject to strict data protection and data confidentiality restrictions. The ethical and legal barriers associated with data sharing and matching are highlighted. Guidance is given on the requirements for confidentiality to protect the identity of the individuals or organizations, including the provision of a controlled, secure physical environment for the computer processing system and the data files.

The methodology for record linkage is presented in a non-technical way so that it is accessible equally to novice and experienced government statisticians. The report offers practical advice on setting up a new matching application and the resolution of the issues to be addressed. The design questions that should be considered when developing record=linkage systems for specific applications are discussed, and suitable methods presented.

http://www.statistics.gov.uk/downloads/theme_other/GSSMethodology_No_25_v2.pdf

3.8     Gomatam, S., Carter, R., Ariet, M., & Mitchell, G. (2002). An empirical comparison of record linkage procedures. Statistics in Medicine, 21(10), 1485–1496.

We consider the problem of record linkage in the situation where we have only non-unique identifiers, like names, sex, race, etc., as common identifiers in databases to be linked. For such situations, much work on probabilistic methods of record linkage can be found in the statistical literature. However, although many groups undoubtedly still use deterministic procedures, not much literature is available on deterministic strategies. Furthermore, there appears to exist almost no documentation on the comparison of results for the two strategies. In this work, we compare a stepwise deterministic linkage strategy with a probabilistic strategy, as implemented in AUTOMATCH, for a situation in which the truth is known. The comparison was carried out on a linkage between medical records from the Regional Perinatal Intensive Care Centers database and educational records from the Florida Department of Education. Social security numbers, available in both databases, were used to decide the true status of each record pair after matching. Match rates and error rates for the two strategies are compared and a discussion of their similarities and differences and strengths and weaknesses is presented.

http://www3.interscience.wiley.com/journal/93516862/abstract?CRETRY=1&SRETRY=0

3.9     Grannis, S. J., Overhage, J. M., & McDonald, C. J. (2002). Analysis of identifier performance using a deterministic linkage algorithm. Proceedings of the 2002 American Medical Informatics Association Symposium, 305–309.

As part of developing a record linkage algorithm using de-identified patient data, we analyzed the performance of several demographic variables for making linkages between patient registry records from two hospital registries and the Social Security Death Master File. We analyzed samples from each registry totaling 6,000 record-pairs to establish a linkage gold-standard. Using Social Security Number

as the exclusive linkage variable resulted in substantial linkage error rates of 4.7% and 9.2%. The best single variable combination for finding links was Social Security Number, phonetically compressed first name, birth month, and gender. This found 87% and 88% of the links without any false links. We achieved sensitivities of 90% to 92% while maintaining 100% specificity using combinations of Social Security Number, gender, name, and birth date fields. This represents an accurate method for linking patient records to death data and is the basis for a more generalized de-identified linkage algorithm.

http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=2244404&blobtype=pdf

3.10    Grannis, S. J., Overhage, J. M., Hui, S., & McDonald, C. J. (2003). Analysis of a probabilistic record linkage technique without human review. Proceedings of the 2003 American Medical Informatics Association Symposium, 259–263.

We previously developed a deterministic record linkage algorithm demonstrating sensitivities approaching 90% while maintaining 100% specificity. Substantially better performance has been reported using probabilistic linkage techniques; however, such methods often incorporate human review into the process. To avoid human review, we employed an estimator function using the Expectation Maximization (EM) algorithm to establish a single true-link threshold. We compared the unsupervised probabilistic results against the manually reviewed gold standard for two hospital registries, as well against our previous deterministic results. At an estimated specificity of 99.95%, actual specificities were 99.43% and 99.42% for registries A and B, respectively. At an estimated sensitivity of 99.95%, actual sensitivities were 99.19% and 98.99% for registries A and B, respectively. The EM algorithm estimated linkage parameters with acceptable accuracy, and was an improvement over the deterministic algorithm. Such a methodology may be used where record linkage is required, but human intervention is not possible or practical.

http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=1479910&blobtype=pdf

3.11    Kelman, C. (2000). The Australian National Death Index: An assessment of accuracy. Australian and New Zealand Journal of Public Health, 24(2), 201–203.

Objective: The Australian National Death Index (NDI) provides a comprehensive and accessible source of mortality information for epidemiological research. Use of the index requires a probabilistic matching process that inevitably results in some inaccuracy. In this paper, accuracy is assessed.

Methods: Results of a matching process against the NDI performed by the Australian Institute of Health and Welfare in Canberra were compared with information provided by the Medical Device Outcomes study cohort and their families (n=2,990). Indices of accuracy for the NDI were calculated.

Results: For this particular study, the NDI has sensitivity of 88.8% (84.9–92.8%) and specificity of 98.2% (97.4–98.7%).

Conclusions and implications: The relatively low sensitivity is of some concern to those using the NDI for health outcomes research. The importance of such a national database is evident; however, to improve accuracy the introduction of a national unique patient identifier is necessary.

http://www3.interscience.wiley.com/journal/119012667/issue

3.12    Li, B., Quan, H., Fong, A., & Lu, M. (2006). Assessing record linkage between health care and vital statistics databases using deterministic methods. BMC Health Services Research, 6, 48.

Background: We assessed the linkage and correct linkage rate using deterministic record linkage among three commonly used Canadian databases, namely, the population registry, hospital discharge data, and Vital Statistics registry.

Methods: Three combinations of four personal identifiers (surname, first name, sex, and date of birth) were used to determine the optimal combination. The correct linkage rate was assessed using an unique personal health number available in all three databases.

Results: Among the three combinations, the combination of surname, sex, and date of birth had the highest linkage rate of 88.0% and 93.1%, and the second highest correct linkage rate of 96.9% and 98.9% between the population registry and Vital Statistics registry, and between the hospital discharge data and Vital Statistics registry in 2001, respectively.

Conclusion: Our findings suggest that the combination of surname, sex, and date of birth appears to be optimal using deterministic linkage. The linkage and correct linkage rates appear to vary by age and the type of database, but not by sex.

http://www.biomedcentral.com/content/pdf/1472-6963-6-48.pdf


3.13   Lui, S., & Wen, S. W. (1999). Development of record linkage of hospital discharge data for the study of neonatal readmission. Chronic Diseases in Canada, 20(2), 77–81.

Computerized record linkage has been used increasingly in epidemiologic studies. We developed a multi-stage, deterministic matching algorithm using various combinations of key variables. Then, from the records for March 1, 1993, to March 31, 1996, contained in the discharge abstract database of the Canadian Institute for Health Information (CIHI), we examined the relation between length of hospital stay at birth and neonatal readmission. A combined use of province/territory of occurrence, 6-digit postal code of residence, date of birth and sex (step 1) matched 88.5% of 26,629 eligible neonatal readmission records with their birth records. Additional use of institution code and chart number or health card number combined with date of birth and sex (step 2 and step 3) increased the matching rate to 93.0%. Compared with the gold standard, step 1 correctly matched 94.4% of the records. We conclude that this deterministic matching algorithm is a feasible and convenient approach to data linkage for the study of neonatal readmission. The linkage strategy may also be helpful in epidemiologic studies of other short-term events.

http://www.phac-aspc.gc.ca/publicat/cdic-mcc/20-2/c_e.html


3.14   Lynch, B. T., & Arends, W. L. (1977). Selection of a surname coding procedure for the SRS record linkage system. Washington, DC: U.S. Department of Agriculture, Sample Survey Research Branch, Research Division.

The Statistical Reporting Service (SRS) is developing a record linkage system to create a master list sampling frame of farm operators in each State. All samples for probability and non-probability surveys conducted by each State Statistical Office (SSO) will be selected from this list. This system uses a probability model which incorporates some of the theoretical concepts developed by Ivan P. Fellegi and Alan B. Sunter. Implicit in the development of their theory is the assumption that if two files are linked then all possible comparisons of all the records of both files will be attempted. However, SRS is really dealing with the "one super file" unduplication problem. That is, different files have been combined into one composite file. The ideal situation in this case is still to make all possible pairwise comparisons. It is clear that even for medium-sized files the number of comparisons under this assumption would be very large. Some technique has to be used to reduce these comparisons to a more manageable number. The

primary objective of this research project was to select the best surname coding technique that could be used to create linkage blocks for the SRS System and thereby reduce the number of comparisons.

http://www.nass.usda.gov/research/reports/Internet_Yield/reportsxyield.html

3.15   Machado, C. J. (2004). A literature review of record linkage procedures focusing on infant health outcomes. Cadernos de Saude Publica, 20(2), 362–371.

Record linkage is a powerful tool in assembling information from different data sources and has been used by a number of public health researchers. In this review, we provide an overview of the record linkage methodologies, focusing particularly on probabilistic record linkage. We then stress the purposes and research applications of linking records by focusing on studies of infant health outcomes based on large data sets, and provide a critical review of the studies in Brazil.

http://www.scielosp.org/pdf/csp/v20n2/03.pdf

3.16   Machado, C. J., & Hill, K. (2004). Probabilistic record linkage and an automated procedure to minimize the undecided-matched pair problem. Cadernos de Saude Publica, 20(4), 915–925.

Probabilistic record linkage allows the assembling of information from different data sources. We present a procedure when a one-to-one relationship between records in different files is expected but not found. Data were births and infant deaths, 1998–birth cohort, city of São Paulo, Brazil. Pairs for which a one-to-one relationship was obtained and a best link was found with the highest weight were taken as unequivocally matched pairs and provided information to decide on the remaining pairs. For these, an expected relationship between differences in dates of death and birth registration was found, and places of birth and death registration for neonatal deaths were likely to be the same. Such evidence was used to solve for the remaining pairs. We reduced the number of non-uniquely matched records and of uncertain matches, and increased the number of uniquely matched pairs from 2,249 to 2,827. Future research using record linkage should use strategies from first record linkage runs before a full clerical review (the standard procedure under uncertainty) to efficiently retrieve matches.

http://www.scielo.br/pdf/csp/v20n4/05.pdf

3.17   Markle Foundation and Robert Wood Johnson Foundation. (2005). Linking healthcare information: Proposed methods for improving care and protecting privacy. New York: Markle Foundation.

The linking of vital information as patients receive care from a fragmented healthcare system is a problem that has consistently plagued interoperability efforts in healthcare. This document outlines a strategy for linking patient information across multiple sites of care, developed by the Working Group on Accurately Linking Information for Healthcare Quality and Safety, a part of the Connecting for Health effort sponsored by the Markle Foundation and the Robert Wood Johnson Foundation.

http://www.connectingforhealth.org/assets/reports/linking_report_2_2005.pdf

3.18   Miller, P. L., Frawley, S. J., & Sayward, F. G. (2000). IMM/Scrub: A domain-specific tool for the deduplication of vaccination history records in childhood immunization registries. Computers and Biomedical Research, 33(2), 126–143.

IMM/Scrub is a pilot tool developed to assist in the deduplication of vaccination history records in

childhood immunization registries. This problem is complicated by a number of factors including that fact that: (1) some doses are numbered and some are not, (2) doses may have different dose numbers, (3) doses may specify different preparations within a vaccine series, (4) one dose may indicate a combination vaccine and the other dose may specify one component of that combination, (5) two doses may have slightly different dates, and (6) combinations of any of these problems may occur together. IMM/Scrub is designed to help detect 10 different types of vaccination dose duplicates and also allows the user to specify flexibly the conditions in which a duplicate dose might be automatically eliminated. In addition, IMM/Scrub is linked to the IMM/Serve immunization forecasting program, which can provide additional assistance in the data cleaning process. The paper describes (1) the design of the current pilot implementation of IMM/Scrub, (2) the lessons learned during its implementation, and (3) our preliminary experience applying it to data from three immunization databases, from a state, a metropolitan area, and an academic medical center.

http://www.sciencedirect.com/science/journal/00104809

3.19    Newcombe, H. B., Kennedy, J. M., Axford, S. J., & James, A. P. (1959). Automatic linkage of vital records. Science, 130(3381), 954–959.

The high cost of searching manually for large numbers of single documents among vast accumulations of files has hampered the compilation and use of routinely recorded facts about individuals to relate successive events in their lives. It is obvious that the searching could be mechanized, but as yet there has been no clear demonstration that machines can carry out the record linkages rapidly enough, cheaply enough, and with sufficient accuracy to make this practicable.

Our own studies were started as part of a plan to look for possible differentials of family fertility in relation to the presence or absence of hereditary disease. The first step has been the development of a method for linking birth records to marriage records automatically with a Datatron 205 computer. For this purpose use has been made of the records of births which occurred in the Canadian province of British Columbia during the year 1955 (34,138 births) and of the marriages which took place in the same province over the 10-year period 1946-55 (114,471 marriages). An intensive study of the various sources of error in the automatic-linkage procedure was carried out on approximately one-fifth of these files. This paper describes the methods used, and the technical problems and errors encountered.

http://www.sciencemag.org/cgi/content/citation/130/3381/954

3.20    Newman, T. B., & Brown, A. N. (1997). Use of commercial record linkage software and vital statistics to identify patient deaths. Journal of the American Medical Informatics Association, 4(3), 233–237.

We evaluated the ability of a microcomputer program (Automatch) to link patient records in our hospital's database (N = 253,836) with mortality files from California (N = 1,312,779) and the U.S. Social Security Administration (N = 13,341,581). We linked 96.5% of 3,448 in-hospital deaths, 99.3% for patients with social security numbers. None of 14,073 patients known to be alive (because they were subsequently admitted) was linked with California deaths, and only 6 (0.1%) of 6,444 were falsely identified as dead in the United States file. For patients with unknown vital status but items in the database likely to be associated with high 3-year mortality rates, we identified death records of 88% of 494 patients with cancer metastatic to the liver, 84% of 164 patients with pancreatic cancer, and 91% of 126 patients with CD4 counts of less than 50. Hospital data can be accurately linked with state and national vital statistics using commercial record linkage software.

http://www.jamia.org/cgi/content/abstract/4/3/233

3.21   Pates, R. D., Scully, K. W., Einbinder, J. S., Merkel, R. L., Stukenborg, G. J., Spraggins, T. A., et al. (2001). Adding value to clinical data by linkage to a public death registry. Studies in Health Technology and Informatics, 84(2), 1384–1388.

We describe the methodology and impact of merging detailed statewide mortality data into the master patient index tables of the clinical data repository (CDR) of the University of Virginia Health System (UVAHS). We employ three broadly inclusive linkage passes (designed to result in large numbers of false positives) to match the patients in the CDR to those in the statewide files using the following criteria: a) Social Security Number; b) Patient Last Name and Birth Date; c) Patient Last Name and Patient First Name. The results from these initial matches are refined by calculation and assignment of a total score comprised of partial scores depending on the quality of matching between the various identifiers. In order to validate our scoring algorithm, we used those patients known to have died at UVAHS over the eight year period as an internal control. We conclude that we are able to update our CDR with 97% of the deaths from the state source using this scheme. We illustrate the potential of the resulting system to assist caregivers in identification of at-risk patient groups by description of those patients in the CDR who were found to have committed suicide. We suggest that our approach represents an efficient and inexpensive way to enrich hospital data with important outcomes information.

http://iospress.metapress.com/content/lwxvjh0aw623n8l4


3.22   Porter, E. H., & Winkler, W. E. (1999). Approximate string comparison and its effect on an advanced record linkage system. In Record Linkage Techniques–1997: Proceedings of an International Workshop and Exposition (pp. 190–202). Washington, DC: National Academy Press.

Record linkage, sometimes referred to as information retrieval (Frakes and Baeza-Yates, 1992) is needed for the creation, unduplication, and maintenance of name and address lists. This paper describes string comparators and their effect in a production matching system. Because many lists have typographical errors in more than 20 percent of first names and also in last names, effective methods for dealing with typographical error can greatly improve matching efficacy. The enhanced methods of approximate string comparison deals with typographical variations and scanning errors. The values returned by the string comparator are used in a statistical model for adjusting parameters that are automatically estimated by an expectation-maximization algorithm for latent class, log linear models of the type arising in the Fellegi-Sunter model of record linkage (1969). Overall matching efficacy is further improved by linear assignment algorithm that forces 1-1 matching.

http://www.census.gov/srd/papers/pdf/rr97-2.pdf


3.23   Potosky, A. L., Riley, G. F., Lubitz, J. D., Mentnech, R. M., & Kessler, L G. (1993). Potential for cancer related health services research using a linked Medicare-tumor registry database. Medical Care, 31(8), 732–748.

The National Cancer Institute and the Health Care Financing Administration share a strong research interest in cancer costs, access to cancer prevention and treatment services, and cancer patient outcomes. To develop a database for such research, the two agencies have undertaken a collaborative effort to link Medicare Program data with the Surveillance, Epidemiology, and End Results (SEER) Program database. The SEER Program is a system of 9 population-based tumor registries that collect standardized clinical information on cases diagnosed in separate, geographically defined areas covering approximately 10% of the US population. Using a deterministic matching algorithm, the records of 94%

of SEER registry cases diagnosed at age 65 or older between 1973 and 1989, or more than 610,000 persons, were successfully linked with Medicare claims files. The resulting database, combining clinical characteristics with information on utilization and costs, will permit the investigation of the contribution of various patient and health care setting factors to treatment patterns, costs, and medical outcomes.

http://www.jstor.org/pss/3767064
http://www.jstor.org/pss/3765984

3.24   Rabeneck, L., Menke, T., Simberkoff, M. S., Hartigan, P. M., Dickinson, G. M., Jensen, P. C., et al. (2001). Using the national registry of HIV-infected veterans in research: Lessons for the development of disease registries. Journal of Clinical Epidemiology, 54(12), 1195–1203.

The purpose of this article is to describe the structure, function, applications, and limitations of the national VA HIV Registry, and to discuss how investigators developing disease registries in the future could benefit from our experience. We examined the number of AIDS patients and the number of new patients identified to the registry, by year, through December 1996. We verified data elements against information obtained from the medical records at five VA sites. We encountered missing data and problems with data classification. Lack of a standardized data classification system was a problem, especially for the pharmacy and laboratory files. In using VA's national HIV registry we have learned important lessons, which, if taken into account in the future, could lead to the creation of model disease-specific registries.

http://www.jclinepi.com/article/S0895-4356(01)00397-3/pdf

3.25   Sideli, R. V., & Friedman, C. (1991). Validating patient names in an integrated clinical information system. Proceedings of the Annual Symposium on Computer Applications in Medical Care, 588–592.

Methods for validating patient names during the upload of clinical records are described. Exact string matching, Soundex method and a pattern matching algorithm (LCS method) are described and compared to a manual analysis of 10000 patient name pairs. In addition, the types of spelling and typographical errors that occur in patient names in the pathology database at CPMC are described. The data analysis shows that the LCS method performs better than the other techniques when compared to manual analysis.

http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=2247599&blobtype=pdf

3.26   Van den Brandt, P. A., Schouten, L. J., Goldbohm, R. A., Dorant, E., & Hunen, P. M. (1990). Development of a record linkage protocol for use in the Dutch Cancer Registry for epidemiological research. International Journal of Epidemiology, 19(3), 553–558.

A method has been developed to determine the optimal linkage key for record linkage between the cancer registry and a large-scale prospective cohort study in the Netherlands. The proposed linkage procedure is a two-stage process in which the initial computerized linkage using a particular linkage key is followed by visual inspection with additional information to separate the computer matches into true and false positives. In the determination of the optimal key, both informativeness and susceptibility to error of personal identifiers were taken into account. The performance of the various keys in the linkage was expressed in terms of sensitivity and predictive value of a reported computer match. The key, consisting of date of birth, first four characters of the family name, and gender was the optimal choice, with a sensitivity of 98% and an initial predictive value of a computer match of 98%. When additional

information on migration, place of birth, and first initial was collected in the second stage, it was possible to eliminate the false positives from the reported computer matches without loss of true positives. Thus, the sensitivity remained constant whereas the secondary predictive value of accepted matches was maximized.

http://ije.oxfordjournals.org/cgi/content/abstract/19/3/553


3.27    Whalen, D., Pepitone, A., Graver, L., & Busch, J. D. (2000). Linking client records from substance abuse, mental health and Medicaid state agencies (SAMHSA Publication No. SMA-01-3500). Rockville, MD: Center for Substance Abuse Treatment and Center for Mental Health Services, Substance Abuse and Mental Health Services Administration.

This report describes the concepts behind record linking and the specific application of record linking in building databases integrating information about mental health (MH) and alcohol/drug (AOD) services.

A variety of methods can be employed to link records from different data sources and these methods vary in terms of complexity, efficiency, and accuracy. Simple matching and deterministic methods are useful for certain applications, and while these methods are relatively simple to implement, they can also produce inaccurate results. By contrast, probabilistic linking methods are relatively complex, but tend to produce more accurate results. The theoretical underpinnings of various approaches to record linkage are discussed, along with the relative strengths of each approach.

Probabilistic linking routines were developed for use in combining Medicaid data with MH/AOD agency data for three States. The nature and function of these routines are described in light of the experience gained in processing State data. Results suggest that when compared with other record linkage methods, probabilistic matching produces more links than other methods and that many of these links are missed by other methods. This indicates probabilistic linking routines are more accurate than other routines for matching person-level data.

To facilitate dissemination of these linking routines, the source code used in the linking process is disseminated at no cost via the project web site. Potential applications and extensions of this methodology are discussed and future directions are outlined.

http://csat.samhsa.gov/IDBSE/idb/modules/linking/material/linking.pdf


3.28    Williams, B. C., Demitrack, L. B., & Fries, B. E. (1992). The accuracy of the national death index when personal identifiers other than social security number are used. American Journal of Public Health, 82(8), 1145–1147.

The study analyzed the accuracy of the National Death Index using personal identifiers that include and exclude the Social Security number. Computerized records of the Department of Veterans Affairs were used for comparison. Different combinations of identifiers other than Social Security number correctly identified 83–92% of dead and 92–99% of living persons. These results should prove useful in ascertaining the mortality status of patient populations without Social Security numbers.

http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=1695740&blobtype=pdf


3.29    Winkler, W. E. (1995). Matching and record linkage. In B. G. Cox, et al. (Eds.), Business survey methods (pp. 355–384). New York: J. Wiley.

Record linkage is used in creating a frame, removing duplicates from files, or combining files so that

relationships on two or more data elements from separate files can be studied. Much of the record linkage work in the past has been done manually or via elementary but ad hoc rules. This chapter focuses on computer matching techniques that are based on formal mathematical models subject to testing via statistical and other accepted methods.

http://www.census.gov/srd/papers/pdf/rr93-8.pdf

## 4.0 FURTHER READING – TECHNOLOGY

4.1     Beckwith, B. A., Mahaadevan, R., Balis, U. J., & Kuo, F. (2006). Development and evaluation of an open source software tool for deidentification of pathology reports. BMC Medical Informatics and Decision Making, 6, 12.

Background: Electronic medical records, including pathology reports, are often used for research purposes. Currently, there are few programs freely available to remove identifiers while leaving the remainder of the pathology report text intact. Our goal was to produce an open source, Health Insurance Portability and Accountability Act (HIPAA)-compliant deidentification tool tailored for pathology reports. We designed a three-step process for removing potential identifiers. Each pathology report was reviewed manually before and after deidentification to catalog all identifiers and note those that were not removed.

Results: 1254 (69.7%) of 1800 pathology reports contained identifiers in the body of the report. 3439 (98.3%) of 3499 unique identifiers in the test set were removed. Only 19 HIPAA-specified identifiers were missed. Of 41 non-HIPAA identifiers missed, the majority were partial institutional addresses and ages. There was variation in performance among reports from the three institutions, highlighting the need for site-specific customization, which is easily accomplished with our tool.

Conclusion: We have demonstrated that it is possible to create an open-source deidentification program which performs well on free-text pathology reports.

http://www.biomedcentral.com/content/pdf/1472-6947-6-12.pdf


4.2     Dumortier, J. The European Directive 1999/93/EC on a community framework for electronic signatures. In A. R. Lodder & H. W. Kaspersen (Eds.), eDirectives: Guide to European Union law on e-commerce: Commentary on the directives on distance selling, electronic signatures, electronic commerce. Vol. 14. Law and Electronic Commerce Series (pp. 33–65). London: Kluwer Law International.

The article discusses the European Community directive on electronic signatures and provides a framework within which electronic signatures can be implemented as certification to confirm the identity of a person. It further describes public key cryptography as the technology behind electronic signatures.

http://www.law.kuleuven.ac.be/icri/publications/58The%20European%20Directive%201999.pdf


4.3     Fulcher, J. (2004). The use of patient biometrics in accessing electronic health records. International Journal of Healthcare Technology and Management, 6(1), 20–31.

Access, ownership, and privacy of medical records are fundamental to the success of any real-world telemedicine application. Such considerations are discussed within the context of smart devices, such as smartcards and iKeys. Unique Patient Identifiers need to be defined before such a scheme would receive widespread adoption. The broader community would also need assurance as to compliance with privacy and other similar legislation. It is further suggested that rather than use (random) digit identifiers, patient biometrics would provide a much better access mechanism, in other words comparing freshly captured biometric identifiers with those stored on the smart device. Experiences gained from a field trial involving the use of USB iKeys for remote access of diabetes patient records are reported upon, and

recommendations made for the future adoption of such systems.

http://www.inderscience.com/search/index.php?action=record&rec_id=4822&prevQuery=&ps=10&m=or

4.4    Halamka, J. (2006). Early experiences with positive patient identification. Journal of Healthcare Information Management, 20(1), 25–27.
To accomplish the goal of positive identification of patients, staff, and medications, Beth Israel Deaconess Medical Center investigated two major kinds of technology—bar codes and radio frequency identification. The article describes different use cases and the pros and cons of each technology.

http://www.hctproject.com/documents.asp?d_ID=4145

4.5    Integrating the Health Care Enterprise. (2007). Patient identifier cross-reference HL7 V3 (PIXV3) and patient demographic query HL7 V3 (PDQV3). Supplement of IHE IT Infrastructure Technical Framework.

This supplement provides a new version of the Patient Identifier Cross-Referencing and Patient Demographics Query profiles leveraging HL7 version 3 and SOAP-based web services. The scope of the Patient Identity Feed, the PIX Query, the PIX Update Notification, and the Patient Demographics Query is identical as that for the HL7 v2.5 messages (i.e., same transaction semantics, same message constraints). In this version we are providing more details for implementers of the individual transactions, and we are using the new 2007 DSTU of the HL7 V3 Patient Topic as the basis of the messages in the transaction. The actual changes to the format compared to the previous year are minimal, as the message content only changes the focal class from identified entity to patient.

http://www.ihe.net/Technical_Framework/upload/IHE_ITI_TF_Supplement_PIX_PDQ_HL7v3_TI_2007_08_15.pdf

Updated document:

http://www.ihe.net/Technical_Framework/upload/IHE_ITI_TF_Supplement_PIX_PDQ_-HL7v3_-TI_2008-11-11.pdf

4.6    Kardas, G., & Tunali, E. T. (2006). Design and implementation of a smart card based healthcare information system. Computer Methods and Programs in Biomedicine, 81(1), 66–78.

In this paper, a smart card based healthcare information system is developed. The system uses a smart card for personal identification and transfer of health data and provides data communication via a distributed protocol which is particularly developed for this study. Two smart card software modules are implemented that run on patient and healthcare professional smart cards, respectively. In addition to personal information, general health information about the patient is also loaded onto the patient smart card. Health care providers use their own smart cards to be authenticated on the system and to access data on patient cards. Encryption keys and digital signature keys stored on smart cards of the system are used for secure and authenticated data communication between clients and database servers over distributed object protocol.

http://ege.academia.edu/GeylaniKardas/Papers

http://www.globalhivmeinfo.org/HIS/Annotated%20Bibliography/Article%209a.pdf

4.7     Leonard, D. C., & Pons, A. (in press). Realization of a universal patient identifier for electronic medical records through biometric technology. IEEE Transactions on Information Technology in Biomedicine.

The advent of universally accessible healthcare data benefits all participants, but one of the outstanding problems that must be addressed is how the creation of a standardized nationwide electronic healthcare record system in the United States would uniquely identify and match a composite of an individual's recorded healthcare information to an identified individual patient out of approximately 300 million people to a 1:1 match. To date, a few solutions to this problem have been proposed that are limited in their effectiveness. We propose the use of biometric technology within our FIRD framework which is a multiphase system whose primary phase is a multilayer composite of these four types of biometric identifiers: 1. Fingerprint; 2. Iris; 3. Retina Scan; 4. DNA. This would allow a patient to have real-time access to all of their recorded healthcare information electronically whenever it is necessary, securely with minimal effort, greater effectiveness, and ease.

http://ieeexplore.ieee.org/xpl/preabsprintf.jsp?arnumber=4534353


4.8     Mathieson, S. (2005). Smartcards. Smart move. Health Service Journal, 115(Suppl. 5972), 12–13.

The article provides an overview of the use of smartcards for healthcare in Europe. The primary use of smartcards has been to speed up reimbursements of medical costs. However, more systems are adding patients' medical information to the smartcards. The article also describes efforts across Europe to integrate smartcard systems.

http://www.ncbi.nlm.nih.gov/pubmed/16171136?dopt=Abstract


4.9     Salkowitz, S. M., & Clyde, S. (2003). De-duplication technology and practices for integrated child-health information systems. Decatur, GA: Public Health Informatics Institute.

Child health integration projects create enterprise-wide, person-centric systems from disparate files with different business rules for identification. Data cleaning activities termed de-duplication are performed to match and merge records appropriately. Projects are challenged to select the most effective d-duplication tools and strategies for their environments.

Interested Connections projects requested this study to research de-duplication software and approaches, perform limited testing and technical analysis, and document the findings in matrices, showing effectiveness, underlying approach, cost and other factors. This report provides a description, analysis and evaluation of de-duplication software based on vendor information and limited testing, documents de-duplication practices of the participating projects, and discusses different approaches and their efficacy.

The study yielded no single best product, but provides a framework to examine alternatives and determine the trade-offs to choose products and strategies that match project requirements. It demonstrates the value of the community of practice and identifies areas for further work.

http://www.phii.org/resources/doc_topics.asp?id=8


4.10    Wilson, S. (2005, May). A novel application of PKI smartcards to anonymise health identifiers. Paper presented at the Australian Computer Emergency Response Team Asia Pacific Information Technology Security Conference, Gold Coast, Australia.

Default thinking about Electronic Health Records (EHRs) and Unique Health Identifiers (UHIDs) has settled on a national numbering scheme, despite the fact that patient privacy can be seriously jeopardised if identifiers ever become linked to individuals' names. A range of generic risk mitigation strategies is envisaged, including strict provider access controls, conservative patient consent provisions, and limiting the amount of personal details recorded for each patient event. Yet none of these measures do anything to control the underlying linkages of identifiers and names, and so a serious gap persists in EHR strategy and architecture. This paper presents a new way to fundamentally anonymise UHIDs through a novel use of public key certificates and smartcards. The design presented here secretes each UHID within an anonymous digital certificate, and links one or more certificates to a smartcard. If an EHR entry is digitally signed via such a certificate, then that entry is directly linked to the UHID, but cannot be linked to the individual's name without having access to the smartcard and the private key it contains. Unique benefits of this approach include strengthened consumer consent controls, efficient off-line identity resolution, reduced reliance on centralised, mission critical identity servers, seamless support for multiple EHRs, and compatibility with a range of smartcard choices available to consumers in the near future.

http://conf.isi.qut.edu.au/auscert/proceedings/2005/wilson05novel.pdf

# 5.0 FURTHER READING – LEGAL, ETHICAL, PRIVACY CONSIDERATIONS

5.1     Australian Institute of Health and Welfare & National Health Information Management Group. (2002). Issues for the use of unique patient identifiers in statistical collections. Canberra: Australian Institute of Health and Welfare.

At its April 2000 meeting, the National Health Information Management Group agreed to accept the following responsibilities in relation to the development and use of unique patient identifiers: (1) identify issues for national minimum data set management raised by proposals for the introduction of unique patient identifiers, (2) draft business rules for the use of unique patient identifiers for linkage for statistical purposes, (3) provide comment and advice on these matters to agencies developing unique patient identifiers, and (4) provide comment and advice on these matters to agencies developing privacy legislation and guidelines.

This paper addresses the first of these four objectives by discussing some of the issues for health and statistical data set management raised by proposals for the introduction of unique patient identifiers (UPIs). The discussion covers UPIs with and without explicitly identifying details such as names and addresses.

http://www.aihw.gov.au/publications/hwi/iuupisc02/iuupisc02.pdf

5.2     Department of Health and Human Services. (2002). Standards for privacy of individually identifiable health information. Final rule. Federal Register, 67(157), 53182–53273.

DHHS modifies certain standards in the rule entitled ''Standards for Privacy of Individually Identifiable Health Information'' (''Privacy Rule''). The Privacy Rule implements the privacy requirements of the Administrative Simplification subtitle of the Health Insurance Portability and Accountability Act of 1996. The purpose of these modifications is to maintain strong protections for the privacy of individually identifiable health information while clarifying certain of the Privacy Rule's provisions, addressing the unintended negative effects of the Privacy Rule on health care quality or access to health care, and relieving unintended administrative burdens created by the Privacy Rule.

http://www.hhs.gov/ocr/privacy/hipaa/understanding/summary/guidanceallsections.pdf

5.3     Goldman, J., & Mulligan, D. (1996). Privacy and health information systems: A guide to protecting patient confidentiality. Washington, DC: Center for Democracy and Technology.

The goal of this report is to help designers of health information systems and policies understand the crucial role privacy plays in our health care system, and in maintaining individual dignity, autonomy, and freedom. By identifying the safeguards that must be built into health information systems and networks, this guidebook seeks to create a standard for incorporating privacy considerations into every aspect of the data collection and distribution process. By protecting the confidentiality of medical data, those who create such systems can ensure that patient privacy is respected and maintained, and that personal information remains accurate, credible, and consistent with the needs of health care providers and patients.

http://www3.interscience.wiley.com/journal/119099433/abstract

5.4     Joint United Nations Programme on HIV/AIDS. (2006, May). Interim guidelines on protecting the
        confidentiality and security of HIV information. Presented at a workshop at a UNAIDS meeting,
        Geneva, Switzerland.

        A three-day workshop was held in Geneva, Switzerland, that was attended by a multidisciplinary group
        of health professionals and community members, including people living with HIV. The workshop's
        aim was to develop draft guidelines on protecting the confidentiality and security of HIV information,
        and to produce a plan to field test them within countries. It involved plenary sessions and small and
        large group work. The main conclusions, recommendations, and next steps are presented in these interim
        guidelines.

        http://data.unaids.org/pub/Manual/2007/confidentiality_security_interim_guidelines_15may2007_en.pdf


5.5     Karp, D. R., Carlin, S., Cook-Deegan, R., Ford, D. E., Geller, G., Glass, D. N., et al. (2008) Ethical and
        practical issues associated with aggregating databases. PLoS Med, 5(9), e190.

        We convened a panel of bioethicists, scientists, and legal experts to analyze the ethical concerns that
        arise when data are shared in aggregated databases and to develop guidelines for aggregating databases.
        Our analysis focused on the aggregated database ImmPort (Immunology Database and Analysis Portal),
        a Web-based resource being developed for the National Institute of Allergy and Infectious Diseases
        (NIAID). Observations about ImmPort should be relevant to other efforts directed at aggregating
        databases.

        http://medicine.plosjournals.org/archive/1549-1676/5/9/pdf/10.1371_journal.pmed.0050190-S.pdf


5.6     Kruse, R. L., Ewigman, B. G., & Tremblay, G. C. (2001). The zipper: A method for using personal
        identifiers to link data while preserving confidentiality. Child Abuse & Neglect, vol. 25(9), 1241–1248.

        Objective: This report describes a method for linking separate confidential data sets that contain
        personal identifying information while preserving required anonymity.

        Methods: Research data were linked with child abuse and neglect (CAN) report data by an independent
        "safe" analyst using an identical set of unique identifier codes assigned to each case in both data sets
        after all personal identifiers had been removed.

        Results: The research team never learned CAN report status of individuals, the state agency never saw
        the research data, and the desired analyses were completed using the merged data set.

        Conclusions: The method was successfully used to merge data from separate sources without divulging
        confidential information.

        http://www.sciencedirect.com/science/journal/01452134


5.7     Mizani, M. A., & Baykal, N. (2007). A software platform to analyse the ethical issues of electronic
        patient privacy policy: The S3P example. Journal of Medical Ethics, 33(12), 695–698.

        Paper-based privacy policies fail to resolve the new changes posed by electronic healthcare. Protecting
        patient privacy through electronic systems has become a serious concern and is the subject of several
        recent studies. The shift toward an electronic privacy policy introduces new ethical challenges that
        cannot be solved merely by technical measures. Structured Patient Privacy Policy (S3P) is a software

tool assuming an automated electronic privacy policy in an electronic healthcare setting. It is designed to simulate different access levels and rights of various professionals involved in healthcare in order to assess the emerging ethical problems. The authors discuss ethical issues concerning electronic patient privacy policies that have become apparent during the development and application of S3P.

http://jme.bmj.com/cgi/content/abstract/33/12/695


5.8     Szolovits, P., & Kohane, I. (1994). Against simple universal health-care identifiers. Journal of the American Medical Informatics Association, 1(4), 316–319.

Medical records are becoming fully computerized. Technical, administrative, and economic forces are pushing toward standardization on a single identifier, such as the Social Security number (SSN), to index all records. Consequently, the privacy and security of our medical histories will be severely compromised. We argue that there are sensible and effective technologic means available to reduce the risks of such compromise, and that it is time to design the characteristics we want in our recordkeeping systems.

http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=116211&blobtype=pdf

5.9     United Kingdom Department of Health, The Caldicott Committee. (1997). Report on the review of patient-identifiable information. London: United Kingdom Department of Health.

In March 1996, guidance on the Protection and Use of Patient Information was published by the Department of Health. This guidance required that when the use of patient information was justified, only the minimum necessary information should be used and it should be anonymised wherever possible. In the light of that requirement, and of the deliberations of a joint DH/BMA working group looking at NHS Information Management and Technology (IM&T) security and confidentiality, the Chief Medical Officer established the Caldicott Committee to review the transfer of all patient-identifiable information from NHS organisations to other NHS or non-NHS bodies for purposes other than direct care, medical research, or, where there is a statutory requirement, to ensure that current practice complies with the departmental guidance. We were asked to examine particular flows of patient information, albeit defined in a broad manner, and to make recommendations following their review and this is what we have done in this report.

http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_4068403


5.10    Upshur, R. E., Morin, B., & Goel, V. (2001). The privacy paradox: Laying Orwell's ghost to rest. Canadian Medical Association Journal, 165(3), 307–309.

For a complex health care system to operate effectively, a balance must be struck between protecting privacy and the need to use individuals' information. The call for explicit consent for the use of health information is intended to respect the autonomy of individuals and recognize their right to self-determination. However, it is unclear whether explicit consent would achieve that intended goal, given that the number of uses of individual data is currently unknown and future uses are unknowable. It is debatable whether individuals wish to give explicit consent every time their health information is accessed or processed. This article presents four approaches to overcome this privacy paradox.

http://www.cmaj.ca/cgi/reprint/165/3/307.pdf

# 6.0 FURTHER READING – IMPLEMENTATION/CASE STUDIES

6.1    Frank, L. (2000). Epidemiology: When an entire country is a cohort. Science, 287(5462), 2398–2399.

The Danish government has compiled nearly 200 databases, some begun in the 1930s, on everything from medical records to socioeconomic data on jobs and salaries. What makes the databases a plum research tool is the fact that they can all be linked by a 10-digit personal identification number, called the CPR, that follows each Dane from cradle to grave. The article describes different studies that have been conducted by Danish researchers using these databases. The article also notes the privacy concerns that have limited scientists' access to certain databases and prevented them from taking full advantage of the wealth of registered information.

http://www.scienceonline.org/cgi/content/summary/287/5462/2398


6.2    Lichtner, V., Wilson, S., & Galliers, J. R. (2008). The challenging nature of patient identifiers: An ethnographic study of patient identification at a London walk-in centre. Health Informatics Journal, 14(2), 131–150.

The correct identification of a patient's health record is the foundation of any safe patient record system. There is no building of a "patient history," no sharing or integration of a patient's data without the retrieval and matching of existing records. Yet there can often be errors in this process and these may remain invisible until a safety incident occurs. This article presents the findings of an ethnographic study of patient identification at a walk-in centre in the UK. We offer a view of patient identifiers as used in practice and show how seemingly simple data, such as a person's name or date of birth, are more complex than they may at first appear and how they potentially pose problems for the use of integrated health records. We further report and discuss a dichotomy between the identifiers needed to access health records and the identifiers used by practitioners in their everyday work.

http://jhi.sagepub.com/cgi/content/abstract/14/2/141


6.3    Pedersen, C. B., Gøtzsche, H., Moller, J. O., & Mortensen, P. B. (2006). The Danish civil registration system: A cohort of eight million persons. Danish Medical Bulletin, 53(4), 441–449.

Background: The Danish Civil Registration System (CRS) was established in 1968, when all persons alive and living in Denmark were registered. Among many other variables, it includes individual information on personal identification number, gender, date of birth, place of birth, place of residence, citizenship, continuously updated information on vital status, and the identity of parents and spouses.

Methods: To evaluate the quality and completeness of the information recorded on persons in the CRS, we considered all persons registered on November 4, 2005, that is, all persons who were alive and resident in Denmark at least one day from April 2, 1968 to November 4, 2005, or in Greenland from May 1, 1972 to November 4, 2005.

Results: A total of 8,176,097 persons were registered. On November 4, 2005, 5,427,687 (66.4%) were alive and resident in Denmark, 56,920 (0.7%) were alive and resident in Greenland, 2,141,373 (26.2%) were dead, 21,160 (0.3%) had disappeared, and 528,957 (6.5%) had emigrated. Among persons born in Denmark in 1960 or later, the CRS contains complete information on maternal identity. Among persons born in Denmark in 1970 or later, the CRS contains complete information on paternal identity. Among women born in Denmark April 1935 or later, the CRS contains complete information on all their

children. Among males born in Denmark April 1945 or later, the CRS contains complete information on all their children. The CRS contains complete information on (a) immigrations and emigrations from 1971 onward, (b) permanent residence in a Danish municipality from 1971 onward, (c) permanent residence in a municipality in Greenland from May 1972 onward, and (d) full address in Denmark from 1977 onward.

Conclusion: Data from the CRS are an important research tool in epidemiological research, which enables Danish researchers to carry out representative population-based studies on, for example, the potential clustering of disease and death in families and the potential association between residence and disease and death.

http://www.danmedbul.dk/Dmb_2006/0406/0406-artikler/DMB3816.pdf


6.4     Rotich, J. K., Hannan, T. J., Smith, F. E., Bii, J., Odero, W. W., Vu, N., et al. (2003). Installing and implementing a computer-based patient record system in sub-Saharan Africa: The Mosoriot medical record system. Journal of the American Medical Informatics Association, 10(4), 295–303.

We implemented an electronic medical record system in a rural Kenyan health center. Visit data are recorded on a paper encounter form, eliminating duplicate documentation in multiple clinic logbooks. Data are entered into an MS-Access database supported by redundant power systems. The system was initiated in February 2001, and 10,000 visit records were entered for 6,190 patients in six months. The authors present a summary of the clinics visited, diagnoses made, drugs prescribed, and tests performed. After system implementation, patient visits were 22% shorter. They spent 58% less time with providers (p = 0.001) and 38% less time waiting (p = 0.06). Clinic personnel spent 50% less time interacting with patients, two-thirds less time interacting with each other, and more time in personal activities. This simple electronic medical record system has bridged the ''digital divide.'' Financial and technical sustainability by Kenyans will be key to its future use and development.

http://jamia.bmj.com/content/10/4/295.extract


6.5     Tang, P. C., Ash, J. S., Bates, D. W., Overhage, J. M., & Sands, D. Z. (2006). Personal health records: Definitions, benefits, and strategies for overcoming barriers to adoption. Journal of the American Medical Informatics Association, 13(2), 121–126.

Recently there has been a remarkable upsurge in activity surrounding the adoption of personal health record (PHR) systems for patients and consumers. The biomedical literature does not yet adequately describe the potential capabilities and utility of PHR systems. In addition, the lack of a proven business case for widespread deployment hinders PHR adoption. In a 2005 working symposium, the American Medical Informatics Association's College of Medical Informatics discussed the issues surrounding personal health record systems and developed recommendations for PHR-promoting activities. Personal health record systems are more than just static repositories for patient data; they combine data, knowledge, and software tools, which help patients to become active participants in their own care. When PHRs are integrated with electronic health record systems, they provide greater benefits than would stand-alone systems for consumers. This paper summarizes the symposium's discussions on PHR systems and provides definitions, system characteristics, technical architectures, benefits, barriers to adoption, and strategies for increasing adoption.

http://www.jamia.org/cgi/reprint/13/2/121


6.6     WEDI Strategic National Implementation Process (SNIP), SNIP Transactions Workgroup, & Health ID

66

Card Sub-Workgroup. (2007). Health identification card implementation guide. Reston, VA: Workgroup for Electronic Data Exchange.

The intent of this implementation guide is to enable automated and interoperable identification using standardized health identification cards. The guide standardizes present practice and brings uniformity of information, appearance, and technology to the more than 100 million cards now issued by health care providers, health plans, government programs, and others.

http://www.wedi.org/snip/public/articles/WEDI-Health-ID-Card-Approved.pdf