CONSIDERATIONS AND GUIDANCE FOR COUNTRIES ADOPTING NATIONAL HEALTH IDENTIFIERS

UNAIDS / JC2640E (English original, June 2014)

Contents

Abbreviations

AIDS	acquired immunodeficiency syndrome
ASTM	American Society for Testing and Materials
CEN	European Committee for Standardization
HIV	Human Immunodeficiency Virus
ISO	International Organization for Standardization
NHID	National Health Identifier
OECD	Organisation for Economic Co-operation and Development
PEPFAR	United States President's Emergency Plan for AIDS Relief
TB	Tuberculosis
UN	United Nations
UNAIDS	Joint United Nations Programme on HIV/AIDS

Acknowledgements

This document was produced thanks to the dedication by staff members from CDC, PEPFAR, UNAIDS and a number of independent consultants including from Regenstrief Institute, Indianapolis, USA.

Comments and queries should be addressed to feedbackhealthid@unaids.org.

Suggested citation: UNAIDS/PEPFAR Considerations and guidance for countries adopting national health identifiers, Geneva 17 April 2014

Executive summary

Many lower- and middle-income countries are in the process of scaling up health services as part of strengthening their health sector to provide services for communicable and non-communicable diseases. Increasing numbers of people in these countries will need to have access to prevention and therapeutic services.

To enhance the effectiveness and efficiency of health systems, different health-care-sector policies need to be linked at the policy level and services need to be integrated at the local, subnational and national level.

The scale-up of health services should be accompanied by an increase in the collection of health information. First, this ensures that all information relevant to delivering optimum services is collected for each individual and that longitudinal personal health information is available within the individual's health record. Second, individual-level health information is needed to monitor and evaluate the effectiveness, efficiency, equity and acceptability of service provision at the facility, regional and national level.

The confidentiality and security of this personally identifiable information has to be protected at all levels of the health-care system.

The existence of national health identifiers (NHIDs) ensures that each patient has one unique identity within the health system. This facilitates the development of longitudinal medical records and allows users of services to be tracked across health-care sectors.

If done within the context of maintaining maximum confidentiality of personally identifiable health information, NHID initiatives have the potential to eliminate the multiple parallel and disconnected patient registration mechanisms found in many countries. The NHID is a key mechanism in the process of harmonizing and collating disparate health records that belong to the same person.

As part of the development of NHIDs, relevant health identity cards may need to be produced. This development may involve advanced technologies for machine reading and link biometric markers to an individual. Stakeholders need to decide whether any biometric markers should be included on health identity cards and which technical solutions are the most appropriate.

Establishing a NHID policy framework, and implementing a system that supports it at the national level for all people, is a complex process that requires strategic planning and coordination among key stakeholders.

The type of NHID to be used should be determined via a national consultative process based on national and international standards and implementation governed via a national standards association. A standard for representing patient demographics is a prerequisite to developing the national identifier.

If a national identification mechanism currently exists for non-health-related government functions, such as social services numbers, then a decision must be made regarding whether to adopt the existing system for health or to introduce a parallel identification system for health. In the trade-off between using a dedicated NHID and reusing a national identification mechanism, key issues include:

- the ability to re-identify a citizen from the national identification number, thereby risking the loss of patient privacy;
- the real and perceived risks of identity fraud and its impact on the health sector;
- the acceptability by certain vulnerable and high-risk sectors of the population of a national identifier for health;
- the cost considerations of a parallel identification system for health.

The NHID needs to be usable for hundreds of years if it is going to be a serious long-term solution. It must allow for three to four generations and population growth, and it must not be reused within at least 200 years of a person's death.

The NHID derives its utility from the ability of any provider to link health services to a unique individual. To achieve this, the identifier must be included in a registry of all identifiers, called the national patient registry, where each identifier is associated with identifying information for each person, including demographic information and a variable number of identifiers, and also supports other essential services in managing patient identity. The registry is an integral part of the overall identification system.

The national patient registry must be populated with data of existing people. These data must come from a variety of sources, including vital statistics databases, points of application for and issuing of the NHID card, and patient service locations. The NHID should be linked to the national vital event registration system, if it exists, containing birth and death records. If the national vital event registration system is not under the governance of the ministry of health, then ministerial coordination and cooperation are essential.

To reduce the risk of duplicate entries in a patient registry, a robust mechanism is needed to support de-duplication and record matching. The mechanism is a critical requirement because the registry is established with historical patient data but will be needed in an ongoing manner to support any scenario that may result in temporary or accidental duplicate registration. Consolidating multiple patient records may require a record-matching mechanism, in which probabilistic record matching is the most appropriate choice.

Significant information and communications infrastructure are needed to support implementation of the NHID. The infrastructure generally requires data to flow between a national host and regional host servers, with sufficient capacity to support daily movement and synchronization of data. The infrastructure can consist of public and private networks.

A well-designed NHID number is free of any personally identifiable information that can be used to identify the individual. In particular, location data such as the place of issue or date of birth must not be part of the identification number.

A well-designed NHID number includes a coding mechanism that facilitates the detection of any errors in transcription. A common mechanism for this is to use a check digit.

A well-designed NHID issued throughout the country can be used at all levels in the country to help identify the source of a particular data item without posing any confidentiality risks.

1. Introduction

Many lower- and middle-income countries are in the process of scaling up health services as part of strengthening their health sector and as part of a broader development agenda. Although in many countries such developments are often influenced by communicable diseases, including human immunodeficiency virus (HIV), tuberculosis (TB) or malaria, the need to improve services for non-communicable diseases is increasingly recognized. As the range and depth of services provided increases, increasing numbers of people will need to be served, especially in areas previously neglected by the local health system.

The scale-up of health services should be accompanied by an increase in the collection of health information. First, this ensures that all information relevant to the services provided is collected for each individual using those services so that longitudinal personal health information can be available over time as part of the development of a longitudinal health record for that individual. Second, individual-level health information is increasingly used to monitor and evaluate the effectiveness, efficiency, equity and acceptability of service provision at the facility, regional and national level *(1)*. Use of individual-level information for the second purpose requires the protection of personally identifiable health information by anonymizing or pseudo-anonymizing the information *(2)*. The use of individual-level information must be balanced carefully with the risks associated with breaching confidentiality.

To enhance the effectiveness and efficiency of health systems, the different sectors of these systems need to be linked at policy levels and integrated at the level of service provision. For example, a woman living with HIV who develops carcinoma of the cervix may need to access relevant cancer services and continue to be followed by her HIV clinic for ongoing HIV management; for this process to be optimum, any element of health data generated in either clinic by the woman needs to be linked to her. This can be facilitated by the existence of a unique health identifier, which will not only simplify the development of a comprehensive longitudinal medical history concerning her management in the HIV clinic, but also ensure that the relevant information from the HIV clinic is available to the clinicians managing her cervical cancer. Furthermore, a unique health identifier will facilitate tracking across other health sectors within the woman's country where she uses services. All this needs to be done within the context of maintaining maximum confidentiality of personally identifiable health information.

This document focuses on the development of unique health identifiers at the country level, referred to as national health identifiers (NHIDs). The purpose of the NHID assigned to an individual is to ensure that the individual can be repeatedly and correctly identified, thus facilitating capture and storage of all information relating to the person's interactions with the health system. Use of NHIDs also ensures availability of this information for monitoring and evaluating health system performance at the facility, regional and national level.

This guidance is written principally to support the process of developing NHIDs within middle- and lower-income countries. Although the rationale to establish NHIDs in the context of strengthening health services is clear, the process to develop a solution is more complex and time-consuming. Several countries have made significant progress

towards developing NHIDs, but many countries have not arrived at a suitable solution. The process of developing and adopting a NHID is not without challenges. It needs to involve leadership, relevant technological infrastructure and resources. Universal solutions are being developed but are unlikely to exactly match local conditions. The NHID represents the most direct path, but the ultimate objective is to achieve accurate identification of people using health services within the context of the country. Countries that have developed a real-world solution for NHIDs have typically used a mixture of methods. Real-world solutions involve trade-offs or compromises to accommodate economic, cultural or other country-specific conditions. Knowledge of the key principles of developing a NHID and the associated health system infrastructure required to support it will help determine how to achieve accurate identification of patients on a national scale.

1.1 Identification of patients in a national health system

1.1.1 Scaling up health services

During the past decade, the focus in many resource-limited countries has been on scaling up HIV services as part of their response to the HIV epidemic. The implementation of HIV service delivery in many settings has taken place at an accelerated pace, with the result that in some countries such services have developed as a vertical system. Although this may have allowed for a rapid emergency response to the country's HIV epidemic, it has become apparent that HIV service provision is intimately linked with other services provided within a country's health-care system.

To sustain the successes achieved to date through the HIV emergency response, the health-care systems of many countries, including both preventive and therapeutic services, need to improve. Such improvements include the development of services, improving linkages at policy levels and integration at service provision levels. This has long been recognized for antenatal services, sexual reproductive health and rights, maternal and child health, TB services, services for acquired immunodeficiency syndrome (AIDS)-defining infections, and cancer services, but attention has recently focused on other non-communicable diseases. Due to the rising numbers of people on antiretroviral therapy *(3)*, many people living with HIV are living longer and increasingly likely to develop comorbid conditions such as cancers, cardiovascular disease, diabetes mellitus and other chronic conditions previously seen only rarely in people living with HIV.

The lessons learned from HIV experiences can be applied to other sectors of a country's health-care system, such as the expansion of health care to support non-communicable disease services. In this context, Member States at the United Nations (UN) High-Level Meeting on HIV/AIDS in June 2011 and the UN High-Level Meeting on Non-Communicable Diseases in September 2011 agreed on ambitious targets to be achieved in both of these areas *(4,5)*.

Coordination within and between health sectors can be impeded if populations in a country are mobile, with increased rural-to-urban migration as people follow employment opportunities to support their

families. If some of these individuals have HIV or other chronic conditions, they will likely use health-care services at multiple locations.

For optimum care, it is essential that care providers have access to relevant information held in a patient's medical record, irrespective of where care is provided. For surveillance and programme monitoring and evaluation, it is critical that any person is counted only once in these systems. In both cases, it is essential that identification of patients is correct and that all data captured in the various sites are attributed accurately to a specific individual. The development of NHIDs is therefore an important part of developing a country's health sector information system.

In resource-limited settings, many clinics and hospitals operate with paper-based health records or registers, often resulting in administrative burden and low-quality data. Each clinical service may issue its own independent patient identifier, which may not be linked to other identifiers within or between institutions. Issuing identifiers to patients manually may be easier in the short term, but it can carry the risk of longer-term coding issues. For instance, by constructing a number from components that indicate service area, year service started, and identification of the issuing facility, complexities are introduced that make it more difficult to use the number for long-term care or to correct mistakes.

To reduce the burden of paper-based documentation and reporting, computerized patient record systems are increasingly used to assist in automating parts of these processes. These systems usually reduce the workload needed to maintain longitudinal records and compile the indicator reports from those records. Without proper planning about how to link paper-based identifiers with electronic record identifiers, however, the need to link patient data across services and multiple health-care locations is challenging and sometimes impossible.

1.1.2 Prior work in multilateral collaboration on personal health information

The issues outlined above have become more prominent as the number of people served by health systems increases and affects both individual care and the health system as a whole. A number of those interconnected areas have been addressed by various international agencies that provide guidance for countries.

A World Health Organization document addressed the issue of linking the monitoring and evaluation systems for HIV, maternal and child health, prevention of mother-to-child transmission of HIV and TB services *(6)*. This guidance links different types of patient information at the facility level. Although not explicitly developing guidance on NHIDs, the report emphasizes the need for a single identifier across services at the facility level. The specifics of how to implement such an identifier are beyond the scope of the current document.

The need for countries to develop NHIDs was addressed more explicitly at a Joint United Nations Programme on HIV/AIDS (UNAIDS)/United States President's Emergency Plan for AIDS Relief (PEPFAR) workshop in 2009 *(7)*. This workshop brought together representatives from a number of countries in the process of developing unique identifiers for health or other services – Botswana, Brazil, Kenya,

Malawi, Thailand, Ukraine and Zambia – and a representative from Denmark, which since 1968 has assigned unique identifiers to its citizens. The participants focused primarily on discussing the benefits and challenges of developing standards for national individual identifiers, and specific designs and their potential implementations for health identifiers. The development of a sound health information system was identified as being important to providing reliable information for planning and managing services for HIV and other chronic diseases.

Another important facet acknowledged during the 2009 workshop was the need for work on NHIDs to be integrated with the development and implementation of national guidelines on protecting the confidentiality and security of personal identifiable health information. An UNAIDS/PEPFAR workshop in 2006 on protecting the confidentiality and security of HIV information focused primarily on HIV information *(2)*, but the guidance from this workshop is relevant to all personally identifiable health information. It discusses the need to protect the confidentiality and security of HIV information as an example of personally identifiable health information. It covers the areas of confidentiality and security of personally identifiable health information held or transported as paper-based or electronic information and the legal framework relating to the patient's right to privacy, the responsibility to protect confidentiality, and the appropriate use of such personally identifiable health information held at local or central data repositories. The workshop proceedings were published as interim guidelines, recognizing that they needed to be expanded to include all personal identifiable health information, that they are likely to change over time as technology changes, and that ultimately countries need to adapt them to their own contexts before adopting and implementing them.

1.2 Scope and limitations

This document provides guidance and considerations for the national implementation of individual identifiers. Although this is a broad topic, this document provides an introduction for a general audience and focuses on the anticipated needs of those who have been charged with developing a strategy and implementation plan for solving the challenge of unique patient identification across the health sector within a specific country. The focus of this document is mainly on the public sector, but it is also relevant to countries where a significant proportion of health care is provided through the private sector; in those countries, however, linkages between public and private health sectors will need to be addressed specifically, and locally appropriate solutions will need to be devised and implemented.

This document provides a starting point, highlighting the many and diverse functional areas that need to be considered to produce sustainable success. Providing detailed functional definitions is not the main aim of this document, but generic and commonly used elements of functional requirements are described. Developing a national patient identification system depends critically on developing the appropriate legal framework to protect patient privacy and corresponding public policy to implement such protections. Interactions between the technical implementation of the identification scheme and legal and policy matters are highlighted; a full treatment of establishing a legal and policy framework is, however, beyond the scope of this document.

Establishing a NHID should be part of a larger country programmatic initiative to develop a national health information infrastructure and standards. Several components of this national infrastructure are a prerequisite to establishing the national identifier, in particular the existence of standard patient demographic data, which in turn may be part of a national data dictionary. The process to agree on collecting such standardized data is not discussed in detail in this document. A country may also have other strategic goals towards strengthening the national health system in ways that have an impact on, or are impacted by, an emerging national health information infrastructure. Two potential activities are particularly relevant. First, the development of a national health information architecture will likely include architectural provisions not only for identification of patients but also for provider identification. Provider identification also requires policy guidance and technical infrastructure support. Second, some countries are moving towards developing universal health coverage initiatives or establishing a health insurance industry. With these considerations, the notion of identification of patients also becomes a key concern in tracking payments and reimbursements. Although these system-level aspects may influence the development of a NHID, they are beyond the scope of this document.

1.3 Outline of this document

The material in this document is organized broadly in the sequence of issues that policy-makers will need to consider:

- Section 2 discusses why identification of patients is needed within the health-care setting.
- Sections 3 and 4 discuss NHID systems and the considerations for NHID cards, respectively. The identifier captures the actual identification information, while the identity card is a physical object held and used by the patient to identify themselves to a health-care provider. We treat these two topics separately, but the interrelationship between the identifier and the card is critical, including when a non-health-specific national unique identifier is used for identity verification within the health system.
- Section 5 discusses the importance of developing national patient registries or data warehouses as the core national-level information system that provides information on the use, cost, outcome and impact of health services *(1)*. Any patient registry or data warehouse is underpinned by whether a nationwide patient identification scheme exists; the latter is a core architectural component in any national health information infrastructure.
- Section 6 discusses national infrastructure concerns in implementing a national identifier.
- Section 7 presents some final considerations and points the reader to the next steps to take in the consideration of a NHID system.
- Five annexes are included to summarize the key concepts and provide examples.
- An annotated bibliography of key publications is provided to guide further reading on the topic.

2. Identification of patients in health-care settings

2.1 Patient identity

Patient identifiers are vital for health-care organizations' day-to-day operations, such as the delivery of care, administrative processes, support services, record-keeping, information management, referrals, follow-up and preventive care. In the continuum of care across any health system, reliable identification of patients is often mandatory for services, such as blood transfusions, invasive testing, surgical procedures and medication administration. Patient identifiers are used routinely for *(8)*:

- coordination of patient care services, such as interacting with other service domains (e.g. laboratories, X-ray departments, dietary services and physical therapy services), communicating orders and results, and requesting services, supplies and consultation;
- clinical documentation and information management, such as collecting and organizing orders, results, procedures and notes on to a patient's chart, and retrieving prior clinical documentation from the patient's chart. The patient's chart may be paper-based or an electronic medical record;
- handling administrative functions, including billing and reimbursement;
- collecting, aggregating and performing analyses on multiple sources of patient data for treatment efficacy, research, statistical reporting and planning.

Identification of patients can play a role at different levels of the health-care system or in a public health or population-level setting. A particular health identifier can be unique at one level and not at the next level.

It is important to be able to identify specifically the levels at which patient identifiers are used; this is called the scope of the identifier. We distinguish six levels of scope where patient identification may be needed:

- The lowest level of scope is in a clinic within a facility. The identifier is unique only within a particular section of the health facility, typically used in a paper record such as a logbook.
- At the facility level, the identifier is typically used for a patient's folder, which is used across the facility.
- Organization-wide scope is limited to functions within the provider organization. This can include both public and private health-care organizations that operate in more than one facility. The current use of patient identifiers by most health-care organizations is at this level.
- Enterprise-wide scope includes multiple provider organizations that provide the same or different types of service. Patient identification is unique within each participating organization of the enterprise.
- Nationwide scope is expanded for use among all public health-care organizations and facilities in a

country. This nationwide identifier can also be used for civil registration, vital event registries, and syndromic or case surveillance activities.

- International scope crosses national boundaries and may involve neighbouring countries. This could be of particular importance in regions where patients may receive care across national boundaries or in research studies defined with international cohorts.

This document is concerned with establishing unique patient identifiers at the national level that will allow accurate identification of every citizen included in the national health system.

Patients may be mobile, visit multiple providers, and be treated by multiple and different types of organization across broad geographical areas. Lack of a unique patient identifier presents significant problems in maintaining continuity of care within and between facilities, and across multiple providers, where access to information from multiple care settings is needed. The retrieval and assembly of relevant patient care information from past episodes of care affect the delivery of care. The unique patient identifier is needed to:

- develop a longitudinal health record of a patient's medical history;
- access and integrate information from different providers and provider computer systems;
- support population-based research and development.

Use of universal patient identification allows:

- the assembly of a longitudinal patient record for more informed and effective treatment and prevention services;
- the ability to accurately de-duplicate aggregate information;
- more accurate data quality assessments and higher-quality data;
- more accurate aggregate indicator reporting for programme management;
- enhanced avoidance of fraud with drug stocks and improved accuracy of research.

Supporting continuity of care is the most compelling reason to uniquely identify patients within facilities and across multiple providers and locations.

Country-wide systems designed to precisely identify individuals and events of interest, as is done in a national surveillance system, can be used to avoid inaccurate statistics caused by patient movement from one clinic to another. For example, without a process to uniquely identify individuals, important events may be counted multiple times, which may artificially inflate the amount of disease or exaggerate the severity of its impact.

2.2 Unique patient identification

Unique patient identification is a key component of a health system and has benefits for both patient care and public health.

The value of a national unique identifier over other methods is its enhanced confidentiality,

since neither the name nor other personally identifying information is used to identify the patient each time information is accessed. The number itself, if designed properly, contains no identifying information.

Current methods for identification of patients almost always involve the use of a medical record number,[1] issued and maintained by the provider organization. Often this number is based on an institutional master patient index and the numbering system is specific to the issuing clinic or hospital. Typically, provider organizations use different numbering systems. Even in situations where the systems are similar across multiple institutions managed by a single organization, typically no efforts are made to ensure identifier uniqueness. As a result, patients may be linked to several medical record numbers, each issued by the clinic or hospital that provided care. These numbers provide unique identification only within the issuing organization. If used within a broader context, it becomes virtually impossible based on the number alone to determine which patients are the same across organizations or locations. In instances where the same medical record number was issued in more than one clinic, facility or organization, different individuals may be erroneously considered to be the same person because they share the same medical record number.

For public health monitoring, reporting aggregate data will likely remain a core need, but aggregate-level reporting can be affected by errors in identification of patients, which can result in double counting. Unique patient identification can provide reliable estimates of the patient identification error rate, which can be used by data managers to improve the accuracy of aggregate reporting. Further, in the absence of NHIDs, attempts are made to use population surveys to provide adjustments to reported aggregated counts. Although this can be successful, it does not provide a high degree of accuracy.

A health-care system may also have a quality-improvement programme. Any such initiative depends critically on accurate service use data, which are improved by accurate patient identification.

2.3 Probabilistic medical record matching for identification of patients

The use of a unique health identifier is not the only means to correctly link a single individual with multiple sets of health information and other relevant sources of information such as vital statistics. One approach to uniquely identifying individuals across facilities, clinics or other sites is to include a set of identifying characteristics within each patient record and then use statistical probability to match the individuals based on these characteristics. In this method, each characteristic is assigned a statistical weight to indicate the degree of confidence in the reliability of the characteristic as an identifier. Then a probability that separate records belong to the same individual is calculated. Those records that fall within an acceptable probability of being the same individual are treated as the same individual, and those that fall outside the range are treated as distinct individuals. Electronic medical records can

1 This may be an alphanumeric sequence, but the term number is used for simplicity.

assist this process by automatically displaying records with calculated matching probabilities above a certain score, which the electronic medical record user then manually defines as either a true match or a non-match. The set of characteristics and their associated weights should be specific to the cultural context of the country. Examples of sets of identifying characteristics are "name, sex, age, race and address" and "sex, age and tribal affiliation".

Probabilistic record matching can provide challenging algorithmic problems that are subject to accuracy and precision considerations. For retrospective use on existing records from multiple sources, probabilistic record matching is well established, and considerations of computational complexity of the matching process are not critical. For prospective use in a patient care environment, the matching process has additional challenges. First, an incorrect match may have implications for patient care. Match verification requires both institutional policies and a standard operating procedure to allow auditing of the matched data and should enable errors to be corrected. Second, the matching process must be linked to the patient indexing system and may require significant computational power and a widely available communication infrastructure; considerable resources may be required to implement this online.

In any implementation of a unique identifier system, the design should allow for the need to match a patient against information that is not linked to the identifier, for example historical records or a laboratory system that has not implemented the NHID. Therefore, even comprehensive implementation of an unique identifier should include the ability for record matching.

2.4 Information standards governing identification of patients

The topic of identification of patients in health care has been the subject of standards development by different organizations. In this document we refer to these standards regularly.

The American Society for Testing and Materials (ASTM) has developed two related standards:

- *Standard guide for properties of a universal healthcare identifier (UHID)* (ASTM E-1714-00) describes the properties of the health-care identifier itself – that is, the numerical/character representation of the patient's identity *(9)*.

- *Guide for implementation of a voluntary universal healthcare identification system* (ASTM E-2553-00) describes a scheme for voluntary implementation of the health-care identifier, using an opt-in mechanism controlled by the patient rather than the government *(10)*.

The International Organization for Standardization (ISO) has developed two standards in the Health Informatics series that specifically support patient identification issues:

- *Health informatics: identification of subjects of health care* (ISO/TS 22220:2011) focuses mainly on the key personal information associated with the identifier, biometrics and patient matching to link independent records belonging to the same person *(11)*.

- *Health informatics: patient healthcard data* (ISO 21549) is a multi-part standard that defines data structures held on patients' health cards compliant with the physical dimensions of ID-1 cards, as defined by ISO/IEC 7810 (smartcards). It provides a platform for developing the structure of the patient health-card data. The structure consists of device data, identification data, administrative data, clinical data, electronic prescription, security data and limited/extended clinical data *(12)*.

- *Health informatics: guidance on patient identification and cross-referencing of identities* (CEN/TR 15872) describes the management of patient identification and cross-referencing of identities and provides some practical guidance for addressing implementation of standards, reports, guidelines, methods, etc. (13)

The standards documents described above can be purchased online. Free or reduced-cost access to these standards for lower- and middle-income countries is being negotiated.

3. A NHID system

An unique health identifier with associated implementation and governance systems can scale up to a NHID. The NHID involves a complete system to assign a single unique identifier to each individual in a country and manage that identifier long term. The NHID is then accessible to provider organizations and used to identify the same individual across service providers, regardless of their location and across time. An advantage of the NHID is its enhanced ability to maintain the confidentiality of the individual, as it can be used independently from personal identifying characteristics. If designed and deployed correctly, the identifier can be used without other identifying information, such as name, age and sex.

3.1 Functions of a NHID

The five basic functions that a NHID must support are:

- positive identification of the individual in order to use services or receive care, or for administrative functions;
- identification of information to aid more informed delivery of care and building of a coordinated multidisciplinary patient management record. This record could contain medical data from different practitioners, sites of care and times to form a lifelong view of the patient's medical record and facilitate continuity of care in the future;
- aggregation of information across institutional boundaries for population-based research and planning;
- protection of privacy and confidentiality through accurate and explicit identification of the patient's information and de-identification once the information leaves the primary site through masking (removal) or encryption (encoding);
- reducing health-care operational costs and enhancing the health status of the nation by supporting both automated and manual patient record management, access to care and information sharing.

3.2 Elements of a NHID system

The following seven components comprise the core elements of a NHID system. They must work together in order for the system to perform its functions:

- identifier scheme that consists of alphanumeric characters that do not represent any aspect of the identity of the individual;
- identification information;
- cross-references to local site-specific patient identifiers for existing patient identification numbers;
- mechanisms to hide or encrypt identifiers;
- software to mass-register patients, and accompanying personnel to carry out this task;
- software to search, identify, match, encrypt or in other ways manipulate the underlying information;
- administrative infrastructure, including the central governing authority.

3.3 Implementing a NHID system

Before the implementation of a NHID system, key policy, design and management questions must be answered. The implementation may be the responsibility of one government department but will more likely involve multiple departments. In this section we outline some of the critical determinations to be made.

3.3.1 Roles and responsibilities

Key roles within the government of the national health system should be defined:

- Who in the government is responsible for the definition, methods and issuance of identifiers to people?
- Who has the legal authority and moral and popular support to lead the effort?

3.3.2 Service delivery and governance

The implementation of the NHID will require ongoing management and oversight. Key considerations are:

- Are there sufficient government organizations to oversee this effort?
- Are there sufficient legal protections to protect misuse of patient information and provide accountability for use of the information?
- Can the infrastructure of the system be deployed and maintained in a robust and secure manner?

3.3.3 Constraints and real-world considerations

During planning and design of the NHID, many practical constraints and considerations must be navigated. Some are universal and some are specific to the country context:

- Make an inventory of methods for identifying individuals and their health records already in use in the country, including those implemented by nongovernmental organizations and other entities that are not within the country's public health-care system but for whom the NHID would still be applicable.
- List the needs for the NHID. These needs might include a short-term temporary identification issued to individuals not yet assigned a permanent NHID, family linkages, disease and disaster control, medical statistics, and long-term complete longitudinal health-care records.
- Calculate the cost and resources required: different design options offer different trade-offs to achieve accurate individual identification.
- Develop an implementation plan. The plan should include the process of announcing the initiative to the medical community, government officials, non-governmental organizations and software developers to allow these groups sufficient time to plan for use of the NHID.
- Evaluate the existence and implementation of a country's guidelines on the protection of the confidentiality and security of personally identifiable health information.

- Evaluate the ability of all responsible institutions involved with the NHID project to plan, deploy and oversee ongoing operations of their respective parts of the work.
- Determine geographical scaling and the penetration of the use of the identifier system in the health services and extension to social services.

3.4 An ideal NHID

A simple user-friendly NHID suitable for use by both humans and computers constitutes an ideal choice. International organizations have produced excellent guidelines for determining such NHIDs. There exists a well-developed concept called the sample universal health identification number. The characteristics of such a number as designed by ASTM are outlined in Box 1.

3.5 Calculating the size of the identifier

If the object is to design a number that is sufficiently large for international use, for instance to cross national borders, then the recommendation from ASTM is to use a 32-digit number, which includes 6 digits of check digits and 6 digits of encryption information. Since this discussion has a national scope and encryption standards can be specified in advance, however, we can use a much smaller number. In the case of most countries, a 12-digit number is sufficient for now and centuries into the future, with the last digit being a check digit. The 11-digit patient number portion would yield a number from 1 to 99 999 999 999.

When calculating the minimum required length of a NHID, the following must be considered:

- base population;
- average age of a generation when bearing children;
- average lifespan;
- number of years past death for which a number should be unique (the quiescent period);
- population growth averages.

An example of a NHID size calculation is given in Box 2. The recommended check-digit calculation uses a check-digit formula originally developed for use in universal product codes (see Annex 3).

3.6 Confidentiality and security

When a legal framework for patient privacy has been established and the corresponding public policies are in place, the NHID can become a mechanism to protect the patient. The NHID is an integral part of a patient's information and requires the same confidentiality and security protection as the patient's other information. The NHID can help meet confidentiality requirements by standardizing and strengthening access control, and eliminating the repeated use of other significant amounts of personally identifiable health information every time access to the patient's files occurs.

BOX 1: ASTM CHARACTERISTICS OF AN IDEAL UNIVERSAL HEALTH IDENTIFICATION NUMBER

Functional characteristics

- Accessible: access depends on the establishment of a network infrastructure, the trusted authority, and policies and procedures that support the system.
- Assignable: assignment of the sample universal health identification number or encrypted universal health identifier, regardless of time or place of request, depends on the establishment and functions of a network infrastructure, the trusted authority, and the implementation of policies and procedures that support the system. It also depends on the mechanism to request a sample universal health identification number.
- Identifiable: this depends on the identification information that the trusted authority links to the sample universal health identification number.
- Verifiable: the sample universal health identification number includes a six-digit check code for verification.
- Can be merged to consolidate multiple identifiers that belong to the same individual: the internal data structure of the sample universal health identification number does not directly support merging duplicate or redundant identifiers. They can be linked at the trusted authority.
- Can be split to assign new identifiers to two or more individuals who have been assigned a single identifier in error: there is no inherent support for splitting the sample universal health identification number. New universal health identification numbers can be issued for future use. Splitting for retroactive information must be handled by the trusted authority.

Linkage of lifelong health record

- Can be linked: the sample universal health identification number can function as a data element and support the linkage of health records in both manual and automated environments.
- Can be mapped: with the use of appropriate database system and software, the sample universal health identification number can be used to map currently existing health-care identifiers.

Patient confidentiality and access security

- Content-free: the sample universal health identification number is free of information about the individual.
- Controllable: this depends on the policies and methods adopted by the trusted authority.
- Health-care-focused: the sample universal health identification number is recommended solely for the purpose of health-care application.
- Secure: the sample universal health identification number includes an encrypted universal health identifier that a mechanism for secure operation through the use of encryption and decryption processes. These capabilities depend on the policies and procedures implemented by the trusted authority.
- Dis-identifiable: the encrypted universal health identifier supports multiple encryption schemes offering multiple encrypted universal health identifiers to prevent the identification of the individual being revealed.
- Public: the encrypted universal health identifier's encryption scheme is intended to hide the identity of an individual when linking information. However, public disclosure of a patient identifier without any risk to the privacy and confidentiality of patient information depends on appropriate access security and privacy legislation, similar to other identifiers.

Compatibility with standards and technology

- Based on industry standards: the sample universal health identification number is not based on existing industry standards. It is based on the ASTM *Standard guide for properties of a universal healthcare identifier (9)*.
- Deployable: the sample universal health identification number can be implemented in a variety of technologies, such as scanners and barcode readers.
- Usable: the sample universal health identification number can be implemented in a variety of technologies, such as scanners and barcode readers. The 28-digit identifier presents difficulty for manual computation and transcription. It may be a time-consuming process and subject to human errors.

Design characteristics

- The ASTM guide and the proposed sample universal health identification number do not address the implementation issues and infrastructure requirements.
- Unique: the trusted authority is responsible for the uniqueness of the sample universal health identification number.
- Repository-based: the sample universal health identification number can be stored in a repository.
- Atomic: the sample universal health identification number consists of a 16-digit sequential identifier, a 1-character delimiter, a 6-digit check digit and a 6-digit encryption scheme. It can function as a single compound data element.
- Conciseness: the sample universal health identification number is not concise – it is a 29-character identifier.
- Unambiguous: the sample universal health identification number is unambiguous. It uses numerical characters and a period as a delimiter.
- Permanent: the sample universal health identification number has sufficient capacity to prevent reuse of identifiers.
- Centrally governed: this policy issue is not addressed. The sample universal health identification number requires central administration and is dependent on the establishment and functions of a trusted authority.
- Networked: the sample universal health identification number can be operated on a computer network. It requires establishment of the necessary network and technology infrastructure.
- Longevity: the sample universal health identification number can support patient identification for a foreseeable future.
- Retroactive: the capacity exists for retroactive assignment of the sample universal health identification number to every person in the country.
- Universal: the sample universal health identification number can support patient identification for the entire world population.
- Incremental implementation: the sample universal health identification number can be implemented on an incremental basis. With the development and use of appropriate procedures and establishment of the necessary bidirectional mapping, both the sample universal health identification number and existing patient identifiers can coexist during the time of transition.

Cost and enhanced health status

- The sample universal health identification number has the potential to support the functions of a NHID. Substantial investment of resources, time and effort will be needed for the establishment of administrative and technology infrastructures; the creation of a trusted authority; the design and development of computer software, hardware and communication networks; and the implementation of security measures.

Additional measures to fully and effectively address the privacy concerns should include:

- national governance in the form of legislation;
- appropriate organizational policies and procedures;
- access control and audit trails that allow for detecting and tracking inappropriate access;
- public education through public service announcements, briefings and other communications;
- continuous evaluation and improvement of these protective measures.

A full description of many aspects that need to be covered in order to protect the confidentiality and security of personally identifiable health information is provided in the *Interim Guidelines on Protecting the Confidentiality and Security of HIV Information (2)*.

The systems supporting the use of the NHID require design architectures that keep the identification of personally identifiable health information and its access as two distinct and separate functions within health care. The identifier's role is limited to identifying the patient's record by accessing only the identification segment of the record but not its content. Access control deals with the authentication of the user, for example validation of the user's identifier and password, verification of access privileges, audit trails and physical security. Access control must be supplemented by organizational policies and procedures and national legislation.

The comprehensive design discussed above must be augmented by appropriate ongoing organizational measures to protect the patient's information. A robust access control mechanism including software security, physical access security, encryption protection and an authentication mechanism must be in place to prevent unauthorized access and ensure legitimate access. Training programmes must be in place to ensure all staff with access to personally identifiable health information are aware of their responsibilities and have the necessary skills to perform them consistently and correctly. Security measures include audit trails for tracking inappropriate access and preventive steps against possible misuse. All security measures must be evaluated periodically and improved continuously.

The role of access security is to grant access for authorized use and prevent unauthorized use of data. The role of a NHID is to assist authorized use by accurately identifying the patient and their information.

A NHID alone will not address the patient's identification need. The NHID cannot protect the privacy and confidentiality of the patient's care information or ensure its accurate identification. These functions depend on security measures such as role-based access security, secure communications and appropriate technology infrastructure. Although most of the ASTM characteristics listed in Box 2 deal with compliance by the issuing authority, health-care information is created, maintained, accessed and used by health-care organizations. Positive identification of individuals and proper controls on access to their information are required at the health-care sites.

BOX 2: EXAMPLE OF NHID SIZE CALCULATION

We calculated the size of the NHID in a country where the population is estimated to grow by 20% in the years 2000–2050, and assumed the same growth pattern for the following 50 years (total 40% growth). The current population is 100 million, so the projected population in 100 years is 140 million. The average age of mothers is 22 years and the average lifespan is 66 years.

To identify all people from birth to death and include a quiescent period after death of 200 years before the number can be reissued, we would need a patient identifier with a capacity for approximately 660 million people. This 9-digit number would fit easily into a 12-digit NHID. This would also allow the final digit to be a check digit, and allow for two orders of magnitude, including extra space to ensure the number is sufficient for the future.

A more robust check code that uses more than one check digit would increase the length of the actual identifier but add the ability to detect and correct for a wider range of errors.

Therefore, the major threat to the privacy of patient care information occurs at the user end where the information resides rather than at the issuing end. Appropriate control and security are therefore required both at the point of issue of NHID, such as a central authority, and at the point of use, such as a provider organization.

Encryption ensures storage and communication in a secure format. Only authorized users can decrypt the encrypted identifier. Encryption may be used when the data are in transit, with information crossing communications lines, or at rest with information stored in systems.

Critical functional elements such as access control, identification information, and administrative and technology infrastructures are independent of the scheme used to construct, or the value assigned to, the NHID. The following measures should be implemented by all organizations that generate, access or use personally identifiable health information:

- access protection;
- user authentication;
- audit trails;
- training and education;
- physical security;
- organizational policies and procedures;
- promotion of an organizational culture conducive to protecting privacy;
- appropriate classification of data into identifiable, non-identifiable and non-person-associated, to aid in determining appropriate system security measures;
- built-in computer hardware and software security, in hardware, operating systems, application software, and communication protocols and methods;

- appropriate segregation of computer networks by firewalls into private, semi-private and public networks;
- proper disposal of electronic and paper medical records, by electronic scrubbing of old media using software designed for that purpose and by shredding paper records.

3.7 Properties of a NHID

The identification number should not contain any data such as site identifiers, regional identifiers, or personal identifiers such as initials or dates of birth, as these types of data are likely to be modified over time. Once this happens, all cards issued to date may need interpretation or may become invalid. For example, regional divisions and numbering systems may be changed as populations shift and grow over a 100-year or longer period.

Imbedding personal identifiers within the card or the identification number introduces risks to confidentiality, as this immediately gives away some amount of personally identifiable information. Furthermore, although it is commonly thought that these data will never change, human errors are made in applications and when entering information into data systems. Correcting such errors will then invalidate the identification number issued to the person.

3.8 Establishing the NHID as a data standard

A NHID, when implemented fully, represents a large data source and should be formally accepted as a national data standard. The identifier is uniquely linked to a core set of personally identifiable information for a person, and this set should be standardized and represented in a manner consistent with and informing national practice. The identifier is linked uniquely to a few core variables that are part of the person's demographic information. The ISO technical specification on the identification of subjects of health care provides a standardized set of data elements associated with identification of patients *(11)*. Country-specific additions, such as tribal affiliations, can be standardized further at the national level.

National data and meta-data standards provide a common understanding of data elements for a national core dataset and national reference data. These data standards must be determined in order to exchange data between systems and have an efficient and high-quality result. Although it is technically possible to cross-walk and translate data from one system to another, it is a much simpler, more easily sustained and less costly process to use the same terms in the same way.

When sharing data, it is necessary to establish a common dictionary of terms and values in order for all systems to interpret information in the same manner, and properly handle all data during the processing of screens and reports. For example, storage of a person's name can be done in a number of ways, such as:

- all in one variable, such as "given name, surname" or "surname, given name";
- separating out the given names and surname;

- including a prefix and suffix, in addition to the given names and surname.

Middle names may be one name, several names or only initials. It is also important to consider "also known as" (AKA) and aliases.

The ISO technical specification on the identification of subjects of health care contains provisions to handle names and related information *(11)*.

4. National health card: design and implementation

The NHID issued to a patient will typically also include the issuance of an identifier card. The card can be used by the patient to communicate their NHID to other parties who need to use the NHID information. The card can use a variety of technologies to ensure the confidentiality of the NHID, to minimize the opportunities for fraud, and to aid accurate data entry and recording. The ISO standard for patient health-card data defines a standard for storing health services data on a smartcard when used to access health care *(12)*.

4.1 Scope of a national identification card

For countries that already use unique identifiers in other facets of government, such as civil registration, taxation and insurance, it may be tempting to conclude that the easiest or best option is to adopt a pre-existing identifier for the health sector. It is necessary, however, to distinguish between an apparent identification and an actual identification. For example, driving licences and passports are treated as actual identification, but they are forged relatively easily, to the extent that a considerable black market exists of producing false identification papers. Furthermore, not all users of the health-care system may be eligible for such pieces of identification.

Generally, the larger the scope and use of the identification card, the larger the potential for fraud, with the result that a number of forged cards may be introduced into the system. A national universal identity card could conceivably be used for all identification needs, but caution should be exercised in assuming that a single identifier should be used universally, especially for financial transactions or legal matters. In addition, specific confidentiality considerations are related to a patient's medical data, such as how the data are used and disseminated.

All technological solutions entail risks, which should be identified, evaluated and discussed fully before adoption. For instance, when used as NHID cards, smart-cards do not provide guaranteed identification of an individual. If there are financial, political or other benefits to be made by forging an identity card, then forgeries may exist, and as a result the public could be made less safe by using identity cards.

In addition, universal identifiers can be misused by authorized inside users and unauthorized outside users. There are many documented examples worldwide of abuse of the personally identifiable information stored in databases, such as privacy abuses, overzealous surveillance, and personally identifiable information being used for nefarious purposes. Because of the possibility and history of extensive abuse of universal identifiers, some countries constitutionally or legislatively severely curtail or forbid the use of universal identifiers *(14)*.

Each country must consider carefully the use of the identifier and whether it should be limited to acting only as a health identifier. In addition, to protect the confidentiality of individuals who are assigned identifiers, access to the data associated with the identifier should be restricted and limited to people whose role requires access, for instance a clinician seeking access to electronic medical record data for patient care.

4.2 Issuance of a NHID card

Implementing a NHID and operating an associated card-distribution system requires careful strategic planning. The process includes creating, issuing, accessing, updating and invalidating cards and data. The plan should describe in detail how the system will be sustainably operated and how large segments of the population will be reached within clearly defined time periods. Such a plan should carefully consider the timeframes, methods and points of issuance (Box 3).

At the time of issuance of the NHID and associated card, the individual must present information to verify their identity. For this, additional documentation is needed (see Box 3). Not all documentation has the same value in identity verification. Annex 2 includes a sample scheme for handling document evidence for identity verification.

4.3 Special populations

To avoid potential stigma or discrimination, careful consideration should be paid to how cards and identifiers are deployed. It is essential to engage with people living with HIV and members of key populations and other vulnerable groups, including sex workers, men who have sex with men, people who use drugs and people with disabilities, so potential concerns such as access to cards and care, risk of unlawful access and use by law enforcement agencies and others can be identified and addressed. Being identified as a person living with HIV, a sex worker or a member of a key or vulnerable population can result in stigma or discrimination in some societies. Consequently, a policy for priority issue of NHID cards risks reinforcing that stigma. Although public health use may indicate the need to prioritize issuing of cards to members of key or vulnerable populations, identifying these people may worsen stigma or discrimination.

4.4 Types of NHID cards

A NHID card can be used to correctly identify a person by including some printed identifiers, for example name, sex or birth date, together with a picture. Smart-cards containing a memory chips can be used to improve the identification of the patient by storing biometric identifiers and patient information.

Considerations for deciding on the type of identification card include the following:

- All cards require a properly trained person to operate the equipment needed to generate a card, such as a computer, card printer, camera and laminating machine.
- Plastic or plastic-laminated paper cards are relatively easy and inexpensive to make.
- Smart-cards are considerably more expensive to make. Prices vary depending on the amount of memory on the card, and on the volume and type produced. Costs may come down as production costs are reduced. In addition, smart-card infrastructure requires card-management systems, card-issuing machines, card readers and significant training of card administrators.

BOX 3: CONSIDERATIONS IN THE ISSUANCE OF NHID CARDS

- What is the planned starting date for issuing cards?
- What is the planned timeframe to issue cards to 90% of the target population?
- What are the projected intermediate dates for 30%, 50% and 75% coverage levels?
- What is the total cost of ownership and lifecycle planning?
- What incentives and links to other identification efforts in the country, such as the private sector, insurance providers and civil registration, will be used?
- What will be the earliest issuance the card in the lifecycle of a person, for example during pregnancy, at birth or after birth?
- What are the geographical and age distribution patterns of the population?
- Where are identity cards and other documents issued, such as birth certificates, voting cards, citizen certificates, passports, and driving and other transportation licences? Locations may include post offices, government locations issuing business licences, community centres, hospital and clinic reception points, military induction points and employment application points.
- Should temporary locations be established to issue the cards?
- Will biometric identifiers be used in the medical and possibly other systems to identify the person?
- Will photographs be taken of the applicant?
- Will other biometric information of the applicant be gathered?
- What type of information will be gathered at the point where the identifier is issued?
- Will biometric identifiers be stored on the cards?
- Where else should the cards be issued, such as hospital outpatient and inpatient reception points and clinics?
- Will the cards be issued immediately, or will there be a period of time in which applications are reviewed and approved?
- When the applications are approved, will all locations where identifiers are issued be able to generate the cards?
- How many people will be required to staff the points of application and issuance?
- What periodic training for staff implementing the cards will be required on roles, responsibilities, and security and privacy considerations?
- What training on interviewing patients, issuing cards and support will be required?
- How robust will the information and communication technology system need to be if application or issuance locations are required to be open at all times?
- Who will provide technical support at the application or issuance locations?
- To protect against cards being issued fraudulently, a verification process that considers the following should be formulated:
- Any documents not in the national official language must be translated by a certified translator.
- Depending on the country and varying local conditions in villages and tribal locations, there may be minimal formal documentation in existence to help verify a person's identity. Alternative identification methods may need to be developed to ensure the integrity of the application, such as meeting with a village or tribal elder, village or community health-care worker, religious leader, or other trustworthy source.
- What are the financial and human resources costs of all of the above?

- Smart-cards can provide significant additional value in identifying patients by providing biometric identification data and core patient data, including latest diagnoses, treatments and care.
- Smart-cards contain electronic components that can be damaged by bending, twisting or perforating and are less sturdy than plastic and plastic-laminated cards.

4.5 Biometrics

Biometrics can be helpful in identifying an individual based on one or more physical traits. The most common physical traits used in recognizing an individual include fingerprints, face recognition, photographs and iris scans.

Fingerprint scanners are relatively inexpensive and do not require extensive training to use, but it is important that the equipment being considered is tested in actual use conditions, with some of the people that will be using it. Fingerprint scanners depend on consistent physical characteristics of the person being scanned, so the amount of variance in those characteristics must be evaluated in order to determine the reliability in typical locations and with typical users.

Typical inexpensive fingerprint scanners use a simple optical method of recognizing the ridges in the fingerprints. Forensic-quality optical scanning fingerprint readers have much higher resolution. All optical scanners are affected by skin dryness and hydration, temperatures and a variety of skin conditions; on average they have a 65–85% specificity. They are not very reliable for children under 5 years old, manual labourers and elderly people.

Fingerprint scanners that rely on recognition of vein characteristics in addition to skin ridges are more accurate. Some can be used in more harsh conditions. These readers are currently being used within systems where more accuracy is needed, including automated teller machines and other banking functions to authenticate users.

Iris and face scanners require a powerful desktop computer on which to run programs and store images. Since the volume of data is high, these scanners are less suitable for use in countries that do not have robust high-speed networking to all facilities.

Whatever technology is chosen, biometric readers will need to be available at all patient registration sites. Training is required, but this can be provided at the facility level.

5. The national patient registry

A unique health identifier derives its use from the ability of any provider to link the use of health services to a unique individual. To achieve this, the identifier must be included in a registry of identifiers, where each identifier is associated with identifying information for each person and supports other essential services in managing patient identity. The registry is an integral part of the overall identification system.

5.1 Managing patient identity

Just as a unique health identifier has a scope that ranges from clinic to international level, so the patient registry can also have a scope. At the facility level, the registry is known as the master patient index; at the health enterprise level, the registry is known as the national master patient index. Historically a master patient index at the facility level would likely have been a paper card index, but currently it is now understood to be an electronic system.

At the national level, we refer to a national patient registry, although this may involve a number of linked registries rather than being a single registry. The national registry is a critical information resource at the national level. It will have linkages to the issuing authority and may be linked to civil registration systems or disease registries. Access to the national patient registry should be limited to users in the health sector, and only users with the highest level of trusted authority should govern linkages with other centrally managed national datasets.

The registry provides the information infrastructure to link the core elements of the identifier system together and links the identifier to the rest of the information about the person. Such information must be updated regularly in order to be accurate and usable. The registry should support search and matching services to locate a unique health identity based on partial information.

In more advanced applications, the national patient registry may also be the backbone of a national electronic health record system, linking the actual health record data to the patient's identity in a secure mechanism. The registry provides vital functions in any environment, however, even if the patient's record itself is primarily paper-based. In an application where a health system is undergoing a transition from paper to electronic patient documentation, the search and matching features of the patient registry will be critical.

5.2 Establishing a national registry for health

5.2.1 Architecture choices for a national registry

A nationwide system does not need to have all the data hosted or managed centrally. There are various alternatives to a single data repository, as outlined in ASTM E-1714-00 *(9)*. Each of the architectures mentioned below should also support transitioning from a temporary identification number to a permanent identification number. Such processes require policy support, standard operating procedures, and institutional resources to manage the transition.

Scenario 1: distributed national unique identifier assignment for a single location

A serial number may be assigned in increasing order based on registration. The NHID can either be associated with a provider location code or associated with quasi-unique patient-supplied demographics, to provide the patient with a nationally unique health identifier. This unique NHID is then associated with all subsequent encounter data and the patient's health record is aggregated using this unique identifier.

Since there must be at least site or patient identification information preceded by a non-unique patient number in order to attempt to make the patient number unique, several significant flaws exist in this method. As regional or district populations or provider locations change, boundary changes are a likely occurrence. When the regions or districts are split, combined or modified in any way, and new codes are generated to refer to the new boundaries, the old codes will become incorrect. The question is then raised over whether to reissue the patient number and renumber all existing patient data, or whether to tolerate patient numbers with outdated, incorrect area codes.

If patient identifiers such as birth date or sex are mixed with the number, personally identifiable information is disclosed every time the patient identifier is used. If the birth date were to be recorded incorrectly and a correction required later, the NHID and all existing electronic and paper medical records would have to be changed.

For these reasons, a method that mixes data such as site or personal identifiers, resulting in an identification number that is unique only at one site, is not desirable. Within this context, in federal countries where health care is a state responsibility, state-specific NHIDs may be generated and be useful; however, people may move across state boundaries and state boundaries may change, so ideally the state NHID should have some link to a national NHID.

Scenario 2: centralized national unique identifier assignment for multiple networked locations

The scenario proceeds as for Scenario 1, assuming that all medium and larger facilities are semi-reliably networked and therefore can share data on a semi-real-time basis. Although this approach is highly desirable and helps to avoid duplicating identities, it may not be possible at all locations and may be unreliable or costly.

Scenario 3: centralized national unique identifier assignment for multiple disconnected locations

Identifier uniqueness and ultimately overall system success rely on timely communication between facilities and the central assignment authority. In resource-constrained environments, this communication may take the form of a person travelling to the point of central assignment and prompt sending and receipt of a request electronically, by telephone, on a memory stick or on paper. Paper should be replaced by electronic methods where possible. The process may depend further on centrally produced, pre-distributed, serialized identifier cards, although this should be discouraged as much as possible, as discussed earlier. Furthermore, this approach often works poorly and is undesirable across an entire country.

Scenario 4: hybrid centralized national unique identifier assignment for multiple networked and semi-networked locations

Combining Scenarios 2 and 3 results in a hybrid approach of 70% or more of the patient volume being interconnected and 30% being done in a timely batch and is considered a reasonable compromise in many circumstances. It is important that all sites with medium to high patient volume are interconnected in order for this approach to function reasonably well.[2] It is also desirable that all lower-volume clinics are also networked. Although the data volume between busier clinics in the same region will be higher, the data volume of the remote clinics should be relatively small. A good result may be achievable with daily mobile telephone data-transfer sessions or a memory stick being sent to the closest networked clinic or hospital.

Scenario 5: locally assigned national unique identifier assignment for multiple disconnected locations

All sites are assigned a unique provider-site code called a prefix and a block of sequential numbers. All local systems are configured to use their specific site code and block of numbers. A locally unique serial number is assigned in increasing order based on registration, following the site code. The NHID is provided to the patient in a durable portable form and is associated with all subsequent encounters and data deriving from those encounters.

5.2.2 Linkages between the NHID and other information systems

In establishing the national patient registry, it must be populated with data of existing people. These data come from a variety of sources, such as vital statistics databases, points of application for and issuing of NHID cards, and patient service sites. Both manual entry and automated processing tools will be needed when large numbers of people are to be registered. In many contexts, paper-based systems may also need to play a large role, especially at primary care sites. A systematic process to handle data conflicts, such as for resolving apparent duplicates, must be established. Besides the core data that should be in the registry, a country may wish to include other data. For example, some health systems have programmes oriented around the household, and thus household data would be valuable in the registry. Similarly, the registry may be used to store family relationships.

A NHID will be assigned either in the absence of a pre-existing health-care identifier or in the presence of an existing health facility patient identifier or multiple identifiers. It is likely that any identifiers that exist are unique only at each clinic or facility.

5.3 Integrating the NHID with the health services

If a health-care identifier already exists in a given place of service delivery, then the existing systems, whether paper or electronic, must be altered to accommodate the new

2 This refers to real-time information technology database systems. In most developing countries, encounter/transaction volume and bandwidth considerations will be different because they may be batched, for example overnight or daily.

NHID. This modification may be time-consuming and expensive, since it must include mapping all of the historical data to the new universal health-care identifier. In addition, issuing of NHIDs and capturing and mapping the new unique health identifier may introduce new ways of working. For example, a pre-existing patient identifier that is generated internally and maintained within a local stand-alone electronic health record system may place no burden on the patient. Conversely, a national unique patient identifier may require the patient to present their patient identifier to each distinct health-care entity. Existing technology will probably need to be changed to accommodate the new identifier in existing electronic systems. The specific processes for mapping pre-existing local or site identifiers to a new national patient identifier vary across different health-care settings because existing processes for generating and maintaining health-care identifiers are highly variable.

Assigning and managing national identifiers may either be conducted by a central assigning authority or be distributed and coordinated among many loosely connected or disconnected sites. Although each approach has advantages and disadvantages, real-world constraints may ultimately dictate the development of a particular architecture. For example, unreliable or non-existent network connectivity in resource-constrained settings may dictate that the assignment of unique identifiers be distributed among many poorly connected or disconnected sites. The potential advantages of a central assigning authority include the potential to minimize the likelihood of an unintentional assignment of duplicate patient identifiers and the potential to leverage economies of scale.

Although a national health-care identifier can improve patient identity management, it is not a panacea. Sophisticated matching methods are a necessary component of any robust unique identifier system. Selection of appropriate matching methods is beyond the scope of this document, but additional information can be obtained elsewhere *(15,16)*. From a technical perspective, comprehensive patient-matching methods will still be needed for a variety of functionalities, including the following:

- Patients receive care even when their identifier is missing. In order to retrieve clinical data in the absence of a unique identifier, a matching algorithm using readily available personal identifiers is needed. A unique health-care identifier system ideally introduces no duplicate identifiers; that is, a NHID is assigned to only a single patient, and two or more patients should never share the same unique identifier. Duplicates may arise for a variety of reasons, however, and therefore a system for managing unique health-care identifiers must implement a process for reconciling such duplicates. The process for doing so may follow this general pattern:

    1. Duplicate identifier is detected through a potential variety of automated or manual processes.

    2. Duplicate identifier is deactivated. No further use of the identifier is allowed.

    3. New unique identifiers are assigned to all individuals who shared the single identifier.

- To fully leverage the benefit of a NHID, clinical data collected before implementation of a NHID system must be linked retroactively to the unique identifier. This requires mapping demographics to link old identifying information to the new unique identifier. Depending on the desired result, this may be accomplished minimally by linkage within a national registry or by relabelling all paper records. The large amount of work required to implement comprehensive renumbering of all patient records requires extensive resources and therefore may not be practical.

- Identity theft and the sharing of identifiers pose challenges. It is difficult to detect medical identity theft or sharing of identifiers unless a provider notices a discrepancy in the clinical data – for instance, if some but not other records indicate that a person has diabetes. The use of someone else's information need not be criminal or malicious: it may be as simple as a parent providing their own identification credentials instead of their child's. Systems must include a method for resolving such issues once they are identified because they are difficult to prevent and detect. To prevent the unauthorized use of a unique health-care identifier, the identity of an individual claiming the unique identifier can be confirmed using supplemental identifying information, such as the area where the unique identifier was initially assigned. This information is then matched against other records to ensure an appropriate match.

- Although it may be necessary to assign multiple identifiers to the same person, in many cases a single identifier per person will be sufficient for most needs. Before assigning a new unique identifier, the system should verify that the individual for whom a new identifier is requested does not already possess an identifier.

A suitable written registration system at sites where health care is provided is needed. This system should capture the name, any aliases, local identification numbers in use, other demographic characteristics for registration, and the NHID number for the patient from their identification card. Any information gathered as part of the original application for the NHID should be available once the identification number has been entered. Any program should be able to perform basic reporting and label printing. The program will likely need to be deployed at least at the following locations:

- hospital outpatient and inpatient receptions;

- stand-alone clinics or wards if patients bypass centralized reception points and go directly to those locations.

Where infrastructure and resources at the service delivery site permit, computer programs should be developed that identify and allow editing of information in order to resolve data discrepancies, conflicts or fraud that may arise. These programs need to match any electronic records at the site based on information such as given name, surname, mobile telephone number, age, birth date, sex or country of birth. Searches should be performed for multiple records with similar identification characteristics, and

BOX 4: SAMPLE CALCULATION TO ESTIMATE THE TIME NEEDED TO HANDLE IDENTIFIER APPLICATIONS

If 100 000 applications are received a month, and possible discrepancies are identified in 5% of the applications, then a total of 5000 applications per month need to be evaluated and resolved. Assuming an 8-hour working day, if the average resolution takes 15 minutes, each person can resolve about 32 records per day, or 160 records per week. If the acceptable time to resolve a problem is targeted to be a week, then in an average week 125 applications need to be resolved, which is within the capabilities of 10 people resolving problems.

information may need to be edited. A status field should indicate whether the record is active, inactive, under review, unusable or denied. If the status of the record is unusable or denied, it is helpful to store the reason, such as multiple identities being issued to a single person.

To facilitate this process, there needs to be a national-level group of professionals who are well trained in record matching and identification and data resolution. In addition, there also needs to be a record status of pending resolution issued on any record that appears to be a duplicate, in error or fraudulent in nature. To facilitate timely resolution, a program that shows these pending records should be written to allow sorting by date of application or pending status.

The number of professionals required to perform this function depends on the rate at which applications and registrations occur per month, and the period of time in which resolution should occur. Box 4 provides a simple example to calculate the workload and time this may take.

To determine the amount of time it takes to provide resolution to registry conflicts, a resolution strategy and a set of scripts for resolving data questions must be generated and tested. The amount of time is thus determined, along with other logistical requirements, such as access to a telephone or other communication means required to resolve problems. A tiered response system needs to be set up that can quickly resolve simpler conflicts by authorized professionals. A second tier of supervisory professionals investigates more difficult or complex identification issues.

5.4 Identifying and labelling existing patient records

To use the NHID in conjunction with existing paper records and ensure that manual transcription errors do not occur, paper forms must bear the NHID on the form. Existing forms should be tagged using machine-readable labels wherever possible to reduce the need for manual transcription of the unique identifier. Forms, folders and cards in use need to be reviewed, and a process needs to be put in place to print new records and update existing records. Printer labels can also be used in tracking specimens.

Current real-world processes for registering and identifying patients are highly variable across different health-care settings. Consequently, the processes for assigning unique identifiers will also be variable. Despite this variation, all processes need to include the following actions:

1. A new identifier request is initiated by an authorized local person, such as the patient, a clinician or an administrator.
2. To verify the identity of the person for whom the identifier is to be created, sufficiently identifying patient characteristics are provided, such as name, birth date and sex.
3. Using the supplied identifying information, a local person verifies that no known unique identifier has previously been assigned to this patient.
4. The trusted authority generates a new unique identifier in response to the request from the local person. The local person may also serve as the trusted authority in distributed settings or in settings where the unique identifier is used for a limited local purpose.

6. Nationwide coordination and infrastructure

6.1 National issuing authority

The issuing and maintenance of the NHID and the use of identification information can be handled under a centralized or decentralized administration. This requires a national-level system, often referred to as a patient registry, which allows authorized people to add and edit data in order to build a national master patient index database. The national system should be distributed into the appropriate regional or district sites to enable access to the patient indexes by local providers of care and services. Updates from local sites flow into the national system and are then synchronized with other local versions of the database, as appropriate. The technical issues around distributing and synchronizing these data can be complex and are beyond the scope of this document.

6.2 Ministerial infrastructure

The ability of a country's government to develop, deploy and operate a high-quality unique identifier system requires the following conditions and resources:

- identifying a lead ministry dedicated to champion the work;
- good coordination skills with other ministries as the project evolves;
- an administrative infrastructure with the ability to engage as needed and fully support the effort;
- the financial resources necessary to ensure the success of the project;
- the willingness to support the effort for an initial minimum commitment of 5 years.

Since the use of information held in a national data repository or data warehouse could easily be expanded for purposes for which it was not originally intended, a robust governance infrastructure should be developed when developing the NHID registry.

Projects of this magnitude require a dedicated team of professionals, including a minimum of the following:

- a senior project manager to keep track of the project, implementing the NHID from inception to deployment. Project managers have varying levels of responsibilities and authority;
- a systems analyst responsible for the development of the information system. Systems analysts design and modify systems by turning user requirements into a set of functional specifications that are the blueprint of the system;
- a database designer expert in relational database modelling;
- a policy expert who researches, creates and recommends the adoption of appropriate policy and procedures;
- a senior technology architect with knowledge of information and communications technology and a broad understanding of data architecture and software development;

- information technology technicians at each facility where improvements and enhancements are needed;
- a technical working committee steering group, consisting of representatives of the major stakeholders and technical experts, to steer the progress, evaluate workplans, and make decisions on the project steps and other related matters as needed.

Organizing the work benefits from systematic project management. Planning and operating large information technology infrastructure projects benefit specifically from an information technology service management framework.

6.3 Coordination of national issuance process

When implementing the NHID across all service facilities, the priorities for staging deployment should be as follows:

1. central or regional hospitals – outpatient reception, inpatient reception, patient records (including retrofitting existing patients) and other points of access;
2. satellite clinics of the central hospital;
3. other, smaller hospitals and general health-care facilities;
4. specialist clinics.

6.4 National information and communications infrastructure

The information technology system containing or coordinating the national patient registry generally requires data to flow between national and regional host servers, and possibly additional hosts. For these data to flow throughout various locations within the country, a reliable communications infrastructure is required to support the movement of data. The infrastructure can consist of public or private networks but must have sufficient capacity to support large daily flows of information to achieve data synchronization across a number of hosts.

In order to synchronize dozens or hundreds of systems containing the patient index, a software architect experienced in value-added network design, deployment and use should be consulted. The architect needs to evaluate the requirements of the patient registry within the particular environment of the country and generate the projected communications requirements necessary to keep all systems up to date. These requirements depend on rates of enrolment and ongoing information flows of enrolled users of services.

This requires at a minimum a three-tiered structure in each country: the top tier consists of one or more national systems configured as a stand-alone or clustered systems; the second tier consists of regional hosts; and the third tier consists of the clinic or hospital end-points. Depending on the size of the country, it may be appropriate to have subregional-level hosts such as districts. The number of tiers needed within the country depends on the size of the country, the robustness of the communications

systems, and the volume of the patient registry transactions.

Each of the tiers must be deployed within an environment that has high availability and tight physical and logical access controls. To achieve this, the following need to be available or strengthened, as required:

- climate control and electrical power conditioning to ensure the computers can operate at least at 95% availability;
- physical access controls, including audit trails, to the locations where the computers are housed;
- application access controls;
- system logging and active or periodic log analysis;
- robust national-level pass/fail system and application monitoring;
- support mechanisms to quickly resolve computer system and communications failures and errors.

It should be noted that servers used in data centres may not be the appropriate choice for many countries. Such equipment demands a consistent level of power and air conditioning that may exceed what can be provided. Contemporary hardware intended for use in fan-less, low-power embedded applications may be a better choice for national-level deployments. Careful considerations need to be made to avoid expensive and often unsustainable hardware.

7. Conclusions

Many lower- and middle-income countries are in the process of scaling up health services as part of strengthening their health sector to provide services for communicable and non-communicable diseases. Increasing numbers of people in these countries will need to have access to prevention and therapeutic services. In order to enhance the effectiveness and efficiency of health systems, different health-care-sector policies need to be linked and integrated at the local, subnational and national level to optimize service provision. The scale-up of health services should result in an increase in the amount of health information collected. This is primarily to ensure that all information on the use of services is collected for each individual, resulting in the creation of longitudinal health records; additionally, individual-level health information can be used to monitor and evaluate the effectiveness, efficiency, equity and acceptability of service provision at the facility, regional and national level.

The confidentiality and security of this personally identifiable information must be protected at all levels of the health-care system. The NHID ensures that each patient has one unique identity within the health sector's information system. This facilitates the development of longitudinal medical records and allows users of services to be tracked across health-care facilities and sectors. If done within the context of maintaining maximum confidentiality and security of personally identifiable health information, NHID initiatives have the potential to eliminate the multiple concurrent and disconnected patient registration mechanisms found in some countries. The NHID is a key mechanism in the process of harmonizing and collating disparate health records that belong to the same patient.

One possible way to allow the patient to present their NHID information to a health facility is through the use of a health identification card. Such a card may use advanced technologies for machine reading, but the card itself does not provide guaranteed identity verification, and mistakes and fraud remain possible. To reduce the possibility of errors and fraud, the use of biometric markers for each individual can be considered. Stakeholders need to decide whether biometric markers are appropriate for the country, and which technical solutions are feasible. Establishing a NHID policy framework, and implementing a national-level system that covers all people, is a complex process that requires strategic planning and coordination among key stakeholders. The type of NHID should be determined via a national consultative process based on national and international standards and implementation governed via a national standards association. If a national identification mechanism exists that is used for non-health-related government functions, including social services, then a choice must be made between adopting the existing system for the health sector or introducing a parallel identification system for health. In the trade-off between using a dedicated NHID or reusing a pre-existing system, key issues include the ability to re-identify a citizen from the national identification number, thereby risking the loss of patient privacy; the real and perceived risks of identity fraud and its impact on the health sector; the acceptability of a national identifier for health by certain vulnerable or high-risk sectors of the population; and the cost considerations of a parallel identification system for health. The identification number needs to be usable for hundreds of years if it is going to be a serious long-term solution. The number must allow for three to four

generations and population growth and must not be reused for at least 200 years after a person's death.

The NHID derives its utility from the ability of any provider to link information on the use and cost of health services to a unique individual. To achieve this, the identifier must be included in a national patient registry, where each identifier is associated with actual identifying information for each person, including demographic information and a variable number of identifiers. The registry is an integral part of the overall identification system. A national patient registry must be populated with data of existing people; these data come from a variety of sources, including vital statistics databases, points of application for and issuing of the NHID card, and patient service sites. The NHID should be linked to a national vital statistics registration system, if it exists, that contains birth and death records. If the national vital statistics registration system is not under the governance of the ministry of health, then coordination and cooperation between different ministries are essential. To reduce the risk of duplicate entries in a patient registry, a robust mechanism is needed to support de-duplication and record matching. These are critical requirements as the registry also contains historical patient data needed to support any scenario that may result in temporary or accidental duplicate registration. Consolidating multiple patient records may require a record-matching mechanism, where probabilistic record matching is the most appropriate choice.

Significant information and communications technology infrastructure is needed to support the implementation of the NHID. The infrastructure requires data to flow between a national host and regional host servers, with sufficient capacity to support daily movement and synchronization of data. This infrastructure may consist of public or private networks. A well-designed NHID number must be free of any personally identifiable information that can be used to identify the individual. In particular, location data such as the place of issue or date of birth must not be part of the identification number itself. A well-designed NHID number includes a coding mechanism that facilitates the detection of any errors in transcription. A common mechanism for this is to use a check digit.

A well-designed NHID issued throughout the country can be used at all levels in the country's health sector to help identify the source of particular data, while providing maximum confidentiality and security of these data.

8. References

1. Beck EJ, Santas XM, Delay PR. Why and how to monitor the cost and evaluate the cost-effectiveness of HIV services in countries. AIDS 2008;22(Suppl. 1):S75–85.

2. Interim Guidelines on Protecting the Confidentiality and Security of HIV Information. Geneva: Joint United Nations Programme on HIV/AIDS and United States President's Emergency Plan for AIDS Relief; 2007 (http://data.unaids.org/pub/manual/2007/confidentiality_security_interim_guidelines_15may2007_en.pdf).

3. HIV/AIDS: fact sheet no. 360. Geneva: World Health Organization; 2012 (http://www.who.int/mediacentre/factsheets/fs360/en/index.html).

4. Resolution WHA65.277. Political declaration on HIV and AIDS: intensifying our efforts to eliminate HIV and AIDS. In: 95th plenary meeting, New York, 10 June 2011. Geneva: Joint United Nations Programme on HIV/AIDS; 2011 (http://www.unaids.org/en/media/unaids/contentassets/documents/document/2011/06/20110610_UN_A-RES-65-277_en.pdf).

5. Resolution WHA 66.2. Political declaration of the High-Level Meeting of the General Assembly on the Prevention and Control of Non-communicable Diseases. In: 3rd plenary meeting, New York, 19 September 2011. Geneva: Joint United Nations Programme on HIV/AIDS; 2012 (http://www.who.int/nmh/events/un_ncd_summit2011/political_declaration_en.pdf).

6. Three interlinked patient monitoring systems for HIV care/ART, MCH/PMTCT and TB/HIV: standardized minimum data set and illustrative tools. Geneva: World Health Organization; 2009 (http://www.who.int/hiv/pub/imai/three_patient_monitor/en/).

7. Developing and Using Individual Identifiers for the Provision of Health Services Including HIV: proceedings of a workshop, 24–26 February 2009, Montreux, Switzerland. Geneva: Joint United Nations Programme on HIV/AIDS; 2011 (http://www.unaids.org/en/media/unaids/contentassets/documents/dataanalysis/20110520_Unique_Identifiers_Meeting_Report_Montreux.pdf).

8. Appavu SI. Analysis of Unique Patient Identifier Options Final Report. Washington, DC: United States Department of Health and Human Services; 1997 (http://ncvhs.hhs.gov/app0.htm).

9. Standard Guide for Properties of a Universal Healthcare Identifier (UHID). West Conshohocken, PA: American Society for Testing and Materials; 2007 (ASTM E-1714-00; http://www.astm.org/Standards/E1714.htm).

10. Guide for Implementation of a Voluntary Universal Healthcare Identification System. West Conshohocken, PA: American Society for Testing and Materials; 2007 (ASTM E-2553-00; http://www.astm.org/Standards/E2553.htm).

11. Health Informatics: identification of subjects of health care. Geneva: International Organization for Standardization; 2011 (ISO/TS 22220:2011; http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=59755).

12. Health Informatics: patient healthcard data. Geneva: International Organization for Standardization; 2013 (21549-1:2013) (http://www.iso.org/iso/home/store/catalogue_ics/catalogue_detail_ics.htm?csnumber=60396).

13. Health Informatics - Guidance on patient identification and cross-referencing of identities. Brussels: European Committee for Standardization; 2014 (15872:2014) (http://standards.cen.eu/dyn/www/f?p=204:110:0::::FSP_PROJECT,FSP_ORG_ID:29275,6232&cs=1C3D1A6C66A0A9F91758BF315DD6A0709

14. Hillestad R, Bigelow JH, Chaudhry B, Dreyer P, Greenberg MD, Meili RC, et al. Identity crisis: an examination of the costs and benefits of a unique patient identifier for the U.S. health care system. Santa Monica, CA: RAND Corporation; 2008 (http://www.rand.org/pubs/monographs/MG753.html).

15. Gomatam S, Carter R, Ariet M, Mitchell G. An empirical comparison of record linkage procedures. Stat Med. 2002;21:1485–96.

16. Machado CJ. A literature review of record linkage procedures focusing on infant health outcomes. *Cad Saude Publica*. 2004;20:362–71.

Annex 1
Unique identifier systems: key concepts

As an international standards organization, the American Society for Testing and Materials (ASTM; http://www.astm.org) has developed and published a comprehensive list of conceptual characteristics for unique health-care identifiers. For an in-depth treatment of each conceptual characteristic, see ASTM 1714.3 *(1)*. The degree to which a given unique identifier satisfies these characteristics will influence the degree to which the identifier can support the health-care functions described in Section 1.2. Related conceptual characteristics can be divided further into the following six subgroups:

- Functional characteristics describe global, system-level capabilities enabled by the identifier. These characteristics include being accessible, assignable, identifiable, verifiable, and able to be merged and split. For example, verifiability, or the ability to determine that an identifier is or is not valid, is often accomplished by including a check digit. A check digit is used to detect an error in the identifier and consists of a single digit computed from the other digits in the identifier.

- Longitudinal linkage characteristics highlight the specific ability of an identifier to aggregate data across multiple systems over time. These characteristics include being able to be linked and able to be mapped. For example, a linkable identifier facilitates the ability to aggregate separate health records across isolated systems, while an identifier that can be mapped can be linked to pre-existing identifiers such as a traditional medical record number.

- Confidentiality and security characteristics address the ability of an identifier to protect and preserve patient privacy. These characteristics include being free of content, controllable, health-care-focused, secure, de-identifiable and public. For example, a health-care-focused unique identifier is used exclusively in the context of health care and does not extend to other enterprises. A content-free identifier contains no personally identifying information, and nor does the identifier convey any information other than uniqueness. In addition to supporting security, content-free identifiers avoid a temptation to which many systems designers succumb – that is, using the information contained in the identifier to support system processes. For example, if the patient's year of birth is contained in the identifier, then systems may begin using that information for system functions such as patient registration or age determination. If the identifier format alters or removes information about the year of birth, then the functionality ceases to work. It is important to note that the principle of being content-free stands in direct contrast to biometric identifiers.

- Standards-based characteristics describe the degree to which an identifier complies with existing approaches. These characteristics include being compatible with existing industry standards, deployable and usable. For example, deployable indicates that an identifier should be able to be implemented using different technologies such as smartcards, barcode readers and paper. Many

existing standards-based health-care transactions accommodate unique identifiers, but it is important to confirm that a particular identifier format is compatible with a given health-care transaction standard. For example, the clinical transaction standard HL7 version 2 limits the maximum length of an identifier string for commonly used patient identifier fields to 20 characters (see http://www.hl7.org). Identifier schemes that use more than 20 characters may be incompatible with HL7 version 2.

- Design characteristics highlight properties inherent to either the actual identifier or the system maintaining the identifiers. These characteristics include being unique, repository-based, atomic, concise, unambiguous, permanent, governed, network, long-lasting, retroactive, universal and incremental. For example, a concise identifier is as short as possible to support efficient entry and to minimize recording errors, the time required for use and the storage required. A long-lasting identifier must be sufficiently long to accommodate substantial information content that will scale with populations that increase in size over time.

- The cost–effectiveness characteristic addresses how well a NHID system provides maximum functionality while minimizing deployment and operational costs.

Examples of unique person identifiers

- Serial numbers: creating sequentially increasing integers, with or without leading zeros (0001, 0002, 0003, etc.), is one of the simplest methods for assigning unique numbers and forms the basis for some of the distributed methods described below. In addition to sequentially increasing simple integers, patterns exist for creating identifiers similar to serial numbers involving alphanumeric codes and special formatting. The United States of America social security number is an example of this with embedded information and formatting. Assigning serial numbers requires either a single assignment authority or close coordination among distributed entities.

- Globally or universally unique identifiers: by creating an identifier that has an exceedingly large number of unique values, the probability of the same value being generated twice is infinitesimally small (although not zero). A commonly used globally unique identifier scheme produces 3.4×10^{38} possible values, and the methodology allows unique identifiers to be created locally on any computer (a distributed assignment system). An example of such a globally unique identifier is 3F2504E0-4F89-11D3-9A0C-0305E82C3301. Globally unique identifiers ensure a high degree of uniqueness, but they do not explicitly have a check digit for self-verification and they contain many digits. The long numbers and lack of check digits limit their usefulness for manual data entry applications.

- Quasi-unique personal identifying elements: although not guaranteed to be explicitly unique, a combination of nearly constant personal demographic elements is commonly used to establish unique identity. These elements may include the person's given name, surname, sex, birth date, mother's given name, mother's maiden name, birth father's given name, birth location such as country or country and city at birth, and birth order. These elements are necessary to perform system look-up and identity validation in situations such as when an identifier is lost and the elements are not typically used as the identifier itself.

- Blocked serial number: when health-care organizations are loosely connected or disconnected, identifier assignment must be distributed and coordinated. A blocked serial number combines a centrally determined, nationally unique site code with serial numbers issued by that site to produce a compound code that is guaranteed to be unique. Although the identifiers are guaranteed to be unique, this system may assign multiple identifiers to a single patient; this may occur when a patient receives care at different sites without their previously assigned identifier. One potential security limitation of this approach is that the site type, such as a antiretroviral therapy clinic, may be embedded or implied in the site identifier. If such codes are publicly available, then the identifier might not be fully content-free.

- Biometrics: biometric identifiers, such as voice patterns, fingerprints, iris patterns, facial shapes and vein patterns, are a method of uniquely identifying individuals. The advantage of biometric identifiers is that they are highly specific to an individual, and identity can be verified without resorting to documents or cards that may be lost, stolen, forgotten or altered. Some of the disadvantages of biometric identifiers include the relatively expensive cost of the equipment and training of personnel. Although biometric identifiers generally remain stable over a person's life, there are instances where the identifiers evolve. Voice patterns can change gradually with age or abruptly with illness, fingerprints can degrade or disappear with time, and retinal patterns may change in very young or old people and in people with conditions that affect the eyes, such as diabetes. Additionally, privacy concerns exist regarding the use of biometric identifiers for health-care uses because of the potential for biometrics, particularly fingerprints, to be used by law enforcement agencies; by their nature, biometrics are effectively non-revocable. Furthermore, resource-constrained settings may not be able to feasibly accommodate the technology required to manage biometric identifiers, which may limit their use in developing settings.

- Identifiers with additional functionality: in addition to ensuring uniqueness, health-care identifiers may convey additional features such as privacy and security. ASTM 1714 describes a process to include security information in the identifier to

indicate whether a particular identifier is intended for fully identified use, such as in patient care, or as a private limited-use token, such as for population-level reporting or research. Although incorporating additional features into the identifier may be desirable from a system implementation perspective, it is important to note that the identifier then compromises the principle of being content-free.

Domains such as law, health policy and ethics provide inputs that inform the strategy for deploying unique health-care services identifiers. Technical and process inputs also inform the unique health-care identifier strategy. To maximize the usefulness of a unique health services identifier, the desired informational and functional characteristics of the identifier should be well understood. Furthermore, the preferred system requirements and existing operational constraints that must be accommodated should be identified to ensure the broadest use of the system.

In general, it is never a good idea to insert personal identifiable data in a patient identifier, since the identifier will persist for an extended period of time, while data such as region or district codes can change over time. When the data change, existing patient numbers based on the old codes may be invalidated, requiring a significant amount of effort to remediate the changes.

Reference

1. Standard guide for properties of a universal health care identifier (UHID). West Conshohocken, PA: American Society for Testing and Materials; 2007 (ASTM E-1714-00; http://www.astm.org/Standards/E1714.htm).

Annex 2
Identity verification with supporting documents

Documents are not always of equal value in verifying a person's identity. One approach is to use a scoring system to build a pass/fail test for the identifying documents. Such scoring systems typically contain four or more levels, with decreasing level of credibility:

- Level 1 documents have the highest credibility.
- Level 2 documents are relatively high in credibility but do not have the same level of credibility as level 1.
- Level 3 documents contain less credibility than levels 2 and 3, and documents must be signed or consented by the applicant or by a competent and trusted person witnessing the oral understanding and agreement.
- Level 4 documents have only limited credibility.

In cases where home addressing systems exist nationwide, no documents are recommended to be considered if they do not contain the applicant's name and address.

Level 1 documents include the following:

- records from or issued by health facilities, such as facility identify cards and immunization cards;
- birth certificates and birth cards (original or certified copies);
- citizenship certificates;
- current passports;
- expired passports that have been valid within the past 2 years;
- other identity documents of the same characteristics as a passport, such as diplomatic documents and some documents issued to refugees such as consular reports of birth abroad, certificates of naturalization, and court or marriage or divorce documents that provide proof of a change in name that differs from the primary document presented.

Level 2 documents include the following:

- driving licences issued within the country;
- government photograph cards;
- licences or permits issued under a law of the commonwealth, state or territorial government, such as boat licences;
- identification cards issued to government employees;
- identification cards issued by the commonwealth, state or territorial government as evidence of a person's entitlement to a financial benefit;
- identification cards issued to students at tertiary education institutions.

Level 3 documents include the following:

- documents held by cash dealers giving security over an applicant's property;
- mortgages and other instruments of security held by financial bodies;

- documents from current employers or previous employers within the past 2 years;
- land title records;
- marriage certificates (often useful to determine a woman's maiden name and the names of parents);
- credit cards;
- foreign driving licences;
- government-issued cards, with no photographs, from other nations.

Level 4 documents include the following:

- records from public utilities, such as telephone, water, gas and electricity bills;
- records from financial institutions;
- electoral rolls compiled and verified by the national voting authorities;
- records held under laws other than laws relating to land titles;
- leases and rental agreements;
- rent receipts from licensed real estate agents;
- records of primary, secondary or tertiary education institutions attended within the past 10 years;
- records of membership of professional and trade associations.

An example scoring system could require a minimum of 100 points. Each document from level 1 could be scored at 100 points and therefore be authoritative on its own. Documents from level 2 could be scored at 75 points each, thus requiring two of them or a mix of other documents. Documents from level 3 could score 40 points each, and documents from level 4 could score 20 points each.

It is important to consider all of these documents, and other authoritative documents that could be considered, within the national context and evaluate each type of document and its value in the process. Situations may arise when people do not have the necessary documentation, for example during natural disasters, war or other calamities; contingency procedures must exist to permit both data processing and service provision for these people.

Annex 3
Checksum algorithm

The final digit of a universal product code is a check digit computed as follows *(1)*:

1. Add the digits (up to but not including the check digit) in the odd-numbered positions (first, third, fifth, etc.) and multiply by 3.

2. Add the digits (up to but not including the check digit) in the even-numbered positions (second, fourth, sixth, etc.) to the result of Step 1.

3. Find the remainder of the result of Step 2 divided by 10 (modulo operation) and subtract this from 10 to derive the check digit.

4. If the last digit of the result in Step 2 is 0, then the check digit is 0.

For instance, the universal product code for a box of tissues is 036000241457. The last digit is the check digit 7. If the other numbers are correct, then the check digit calculation must produce 7, as follows.

1. Add the odd-numbered digits: $0 + 6 + 0 + 2 + 1 + 5 = 14$.

2. Multiply by 3: $14 \times 3 = 42$.

3. Add the even-numbered digits: $3 + 0 + 0 + 4 + 4 = 11$.

4. Add the results: $42 + 11 = 53$.

5. To calculate the check digit, take the remainder of (53/10), which is also known as (53 modulo 10), and subtract from 10. The check digit value is 7.

Reference

1. Appavu SI. Analysis of unique patient identifier options final report. Washington, DC: United States Department of Health and Human Services; 1997 (http://ncvhs.hhs.gov/app0.htm).

Annex 4
Sample legal and policy guidance

Establishing a legal or policy framework to govern aspects of data confidentiality and patient privacy is beyond the scope of this document. We provide here an abbreviated sample representation of items that could become part of such legislation or policies.

A health data privacy act for individual patients

The mandate for governance should be achieved through legislation that governs the right to privacy of information for people in a privacy act. As an example of such legislation, a privacy act can include provisions that establish that:

- the definition of personally identifiable information includes written, spoken, electronic and all other forms of information about an identifiable individual;
- collection and use of personally identifiable information by the government can lead to misuse in ways that are specific to government functions, such as using medical data for criminal investigations;
- the national judiciary is granted the authority to review all claims of denial of access to personally identifiable information held by the government, and improper collection, use and disclosure of personally identifiable information;
- the rights of access, correction and notation with respect to personally identifiable information held by a government institution apply to all people, whether citizens of the country or not;
- government institutions can collect personally identifiable information that is reasonable and necessary only for the particular purpose. All collections of data must enumerate the authority under which they are collected, what the data will be used for, whether and with whom the data will be shared, the consequences of not providing the information, and how an individual files a complaint in cases of potential misuse of their personally identifiable information;
- in the case of a disclosure of personally identifiable information that does not occur in a manner consistent with the intended use, there is a corresponding duty on the institution to inform the individual(s) immediately about the disclosure. In addition, a detailed review of the provisions allowing disclosure without consent should be conducted in order to strengthen and clarify any wording, and all users of the data must be instructed about the proper use of the data;
- all data use must be fully disclosed to the individual, with the framework established for this being based on maximizing the transparency of government and ensuring a maximum amount of accountability of government to its people;
- no data can be used for purposes outside the intent disclosed to the individual, and no data matching, linkages or aggregation of the data by external systems is allowed, except as

disclosed in writing to the people to whom the data belong;

- all data system usage will be governed by the principle of least privilege – that is, people who have access to the system are assigned the lowest level of access rights necessary to do their jobs;
- the government institution remains accountable for personally identifiable information where decisions are made to outsource departmental work, and the information is considered to be under the control of the government institution at all times;
- an office of information security or similar government office is responsible for implementing the policies and operational mechanisms to support the privacy act.

An information security policy

Example activities that comprise the activities of an office for information security are:

- establishing standards and requirements for protection of personally identifiable, personal but non-identifying, and non-personal data;
- establishing standards that require all systems to have privacy impact assessments and system security plans approved before operational use of the system;
- providing education for all people using personally identifiable data;
- establishing security standards and requirements for data and data systems;
- establishing security, authentication and access standards and requirements for data systems;
- establishing accreditation standards and requirements for systems;
- establishing data classification standards and requirements for data protection based on such classification;
- establishing appropriate sanctions for non-compliance of requirements or altering, falsifying or concealing any records kept by the government;
- performing periodic audits and reviews of national systems to ensure compliance;
- establishing specific ownership, confidentiality and usage standards for all data, whether collected, matched, aggregated or used for research; as a minimum, patient data are owned by the patient, aggregated or otherwise derived data belong to the government, and the government has rights to use patient data;
- establishing standards for use of citizen data by government officials;
- establishing common business, technology, rules and processes in the key domains of security, privacy and technology for systems that contain sensitive data and for data flows between systems;

- establishing standards and requirements relating to flow of personally identifiable information, including written plans and disclosures about use of the data and confirming that all information leaving the country will be managed through processes that ensure protection of confidentiality in a manner commensurate with national data standards and requirements. The plans should include a description of the personally identifiable information to be shared; the purposes for which the information is being shared and used; a statement of the administrative, technical and physical safeguards required to protect the confidentiality of the information, especially with regard to its use and disclosure; a statement specifying that the sharing of data will cease if the recipient is discovered to be improperly disclosing the shared information; the names, signatures and titles of the officials in both the supplying and receiving institutions; and the date of the agreement;

- establishing a mechanism for release of patient records to the individual to whom the record pertains and requiring notification of any release of data to that individual;

- requiring that all data system users sign an acceptable data use agreement;

- establishing data disposal standards and requirements;

- developing performance indicators to report on the effectiveness of these systems to the executive and legislative arms of government and to ensure that the public is adequately informed of these measures and their implications for informational privacy;

- encouraging the sharing and implementation of best practices in informational privacy management across the national or federal government and promoting the exercise of self-correction, prevention of data loss or misuse, and risk mitigation;

- encouraging and fostering a process of continuous learning and improvement about the codes of ethics that professionals engaged in privacy management need to set for their practice;

- establishing that the office of information security can receive complaints concerning the full array of rights and protections under the privacy act, including complaints regarding inappropriate collection, use or disclosure, failure to maintain up-to-date and accurate data, improper retention or disposal of data, and denial of access or lack of corrections. The director can make recommendations to the government institution, request notification of the actions the institution plans to take, and report the institution's response to the complainant. If the response from the institution is not satisfactory, then the director has no authority to require remediation of deficiencies other than to publicize the conclusion of event in a report. The courts have the sole authority to order a remedy based on the merits of the situation, but they are

precluded from providing guidance on standards, requirements or what constitutes inappropriate collection, use or disclosure of personally identifiable information;

- providing a resource for investigating breaches of confidential information, whether accidental or intentional, and providing recommendations to those responsible on how to better secure this information in the future;
- providing for the ability to levy fines commensurate with the extent of the damages.

Annex 5
Estimating resources for implementing a NHID system

Implementing a NHID system is a large and expensive project in terms of human, time and financial resources. The costs are affected by a number of variables:

- use of a central, distributed or hybrid implementation model;
- expected monthly enrolment rate;
- number of concurrent locations providing enrolment services;
- whether publicly available enrolment places can use existing building space;
- availability and costs of staffing the patient enrolment and other contact points and administrative costs of the national registry;
- geographical distances and population distributions;
- distribution of the NHID registry system;
- which types of identity verification are going to be used, such as photographs or fingerprints;
- whether data and voice communication systems needed to support the effort are adequate;
- whether the scope of the project includes development of national data standards and national data-sharing standards, at least for identification data elements;
- the quality and availability of public roads and transportation systems.

Much of the information that follows is technical in nature because the costs are driven directly by the costs of implementing certain technological models. Therefore, a well-organized evaluation is needed. This evaluation should be an interactive process that includes both information technology professionals with system architectural experience, and management officials. The evaluation should be overseen by a management stakeholder group that identifies appropriate technical resources and forms a technical working group, identifies a set of goals for the technical working group, and sets an acceptable timeline for those goals. By engaging the proper expertise, appropriate solutions and estimated costs can be ascertained with a series of recommendations and discussions of budget and outcome. When coupled with a skilled project manager, a detailed project management plan and good communications with stakeholders, a successful result can be reached.

Implementation models

There are a number of different models for implementing NHIDs, which vary in the number of service points, capacity to scale up and provide national services, complexity and costs. The following three models are examples of the many variations that should be considered.

Heavily centralized model

This model typically leverages existing government services that occur at one central location, such as the capital city. Characteristics of a heavily centralized model include the following:

- Costs are minimized since the model can use enrolment locations that are already providing registration services and possibly are already issuing permits or licences.
- Typically, a single or small number of events induces the citizen to apply for inclusion in the identification system. These events might include birth, reaching a certain age, hospitalization, joining the military, or applying for a driving or marriage licence.
- The model lends itself to centralized administration and issuance of cards.
- Annual costs are typically lower than for other models due to the limited number of people applying on an annual basis, and use of existing infrastructure to receive applications. The costs of processing applications and administrative costs of the national registry are nearly the same per enrolee for all models.

Semi-distributed model

This model is similar to the heavily centralized model, but it leverages existing government services that typically occur at regional or provincial centres. Characteristics of a semi-distributed model include the following:

- Costs are typically somewhat higher than for a heavily centralized model but lower than for a highly distributed model.
- Typically, a single or small number of events induces the citizen to apply for inclusion in the identification system. These events are similar to those for a heavily centralized model.
- Typically, the application points have card printers and can issue cards upon administrative approval.

Highly-distributed model

In this model, NHID administration occurs at the point of health service delivery throughout the country. Characteristics of a highly distributed model include the following:

- There is a much more complex rollout schedule, with a number of teams performing site readiness to meet an aggressive schedule, such as 1–2 years.
- Costs are higher due to the higher number of application and issuance points and the number of teams executing site start-ups.
- If an aggressive schedule is desired, a higher number of rollout teams will be required, and administrative costs will be higher in processing the increased number of applications.
- Many events will typically induce the citizen to apply for inclusion in the identification system, including all of the previously mentioned events plus community enrolment efforts and medical events such as hospitalization or clinic visits.
- A carefully thought-out and available communication system is required, with robust connectivity to at least the district level and probably also the larger site level.

- Rollout costs will be higher due to a larger rollout team, which includes providing equipment, hiring staff and training on the identification system.
- Unless there is a uniform national language, translation will be needed at least for the application forms and possibly also for the data-entry screens.
- This is a more complex NHID registry service model, due to the need to communicate to a wide variety of electronic systems, provide printouts for paper systems, and maintain highly synchronized pools of data to ensure no single point of failure in the national system.

Identifier methodologies and systems

Depending on the personal identifiers used, suitable verification of information must be done during the application approval or denial process, and during delivery of the identification card. Personal descriptors may include the person's photograph, name, height, eye colour, sex and age. It is important to keep in mind that all data stored in the NHID registry and contained on each identification card must be accurate and verified to the highest degree that can reasonably be done. Consistently accurate and timely data require thorough and proper planning and development of clear policies and procedures. In addition, the methods used when the patient presents at the point of service should be clearly established to ensure proper identification occurs.

Identification cards

ISO 7810 is an international standard for the physical characteristics of identification cards. ISO 7810-ID-1 specifies that the cards measure 85.60 mm × 53.98 mm. This is the most common size of bank, credit and debit cards and driving licences.

ISO 7811 is an international standard for recording printed and magnetic data on identification cards. It contains standards for the embossed characters and several specific formats for recording magnetic data.

ISO 7816 is an international standard for identification cards with an embedded chip (smartcards) and electrical connections for the chip.

Use of international standards is highly recommended where applicable and reasonable.

Biometrics

The most common physical traits used in recognizing individuals include height, sex, fingerprints, face (photographs) and iris.

Whatever technology is chosen, biometric readers need to be available at all patient registration and presentation points, which could number in the dozens in a heavily centralized model to thousands in a highly distributed model.

To use this technology successfully, training is required. Training may be accomplished at the facility level to minimize staff disruption or may be located more centrally to maximize the trainer's time.

All prices stated in this document were found in an internet search of various suppliers in

2011 and are used here simply to show the variances of costs.

Fingerprint scanners

Typical inexpensive fingerprint scanners cost US$ 75–150. These scanners use a simple optical method to recognize the ridges in the fingerprints. Higher-quality, higher-resolution forensic optical scanning fingerprint readers cost US$ 400–700. All optical scanners are affected by skin dryness, how much water the person has consumed, low temperatures, and the condition of the skin. They typically have a 65–85% success rate. They do not work well in children under 5 years of age, people whose skin is worn down by performing manual labour, or elderly people.

Fingerprint scanners that rely on sub-dermal characteristics in addition to the skin ridges are more accurate, and some can be used in more harsh conditions. Less expensive units cost US$ 125–300, depending on the quality. Units designed for use in harsher conditions (moderate amounts of water and dust) cost US$ 600–1000. The latter are used in systems where more accuracy is needed, including automated teller machines and other banking functions to authenticate users.

Iris and iris/face scanners

Iris and iris/face scanners cost US$ 900–4000, depending on the accuracy of the equipment. They typically require a software development kit that costs US$ 500–1500. In addition, these scanners require a powerful desktop computer on which to run the programs and store the images. Since the volume of data is much higher with this type of scanner, they are less suitable for use in countries that do not have more current computers and robust high-speed networking to all facilities.

Photography

Photography is the most common and most easily understood way of identifying a person. Use of photographs requires some basic components: a computer, a digital camera, photograph management software, a suitable printer, and the identification card. When using a photograph to aid in the identification of a person, specific procedures and adequate training must be supplied to the people taking the photographs, the people verifying the data and cards, and the people issuing cards to people.

Photographs used to identify people must adequately support use on both identification cards and computer monitors. A commonly used size for a printed photograph on identification cards is 358 × 441 pixels, printed at 350 dots per inch using the RGB format at 24-bit true colour, prepared using JPEG standardized in ISO DIS 10918-1. The resulting printed photograph measures 25.9 × 32 mm.

The photograph is used during the registration administration process and may also be used at the point of service to identify the person by displaying it on a computer monitor. Therefore, the original photograph should have sufficient adequate resolution to support up to a 100 × 152 mm 24-bit visual at 300 dots per inch.

Photographic workstation

Photographs are imaged by a digital camera and stored on a computer. Properly designed computer software should be used to enter the personally identifying information and for digital storage of photographs.

It is important to physically secure workstations, cameras and printers from theft during the day and after hours.

The computer performs the actual tasks as defined by the NHID system. The computer, camera and all related connections should have sufficient speed to process the images in an efficient fashion in order to accommodate the expected daily workflow. The computer should be a dual-core machine with at least 2 GB of memory and a USB interface for the camera. Prices in 2012 were in the range US$ 400–800.

Camera

A digital camera or video camera captures the person's photograph and loads it into software. Digital cameras vary in resolution, speed, simplicity of operation and ruggedness. Regardless of the model chosen, the ability to quickly upload the image from the camera to the computer is an important consideration. Prices in 2012 were in the range US$ 75–400.

Digital printer

A digital printer receives all the text, photographs and images from the NHID system software and prints them directly on to a plastic card. Printer specifications vary, including speed, support of single- or double-side printing, whether it encodes magnetic strips, whether it encodes smartcards, whether it provides lamination, whether it prints in colour, and how close to each edge it prints. Prices in 2012 were in the range US$ 1000–8000.

Communication models

Regardless of the distribution of enrolment and card-distribution sites, the overall communication system should be analysed and a service level decided. The simplest way to develop an overall service level analysis is to list each site and the services provided. Each service entails a certain volume of data per occurrence, and this amount of data should be estimated. For example, each time a new person is enrolled, data from the site performing the enrolment, including the enrolment form, and all biometric information, such as photographs, will need to be communicated to a central registry server. A table of services and data volumes should be constructed.

The next step is to build a matrix of the sites that will need to communicate with each other and decide an acceptable amount of time in hours for this connection to be unavailable (which will cause a service outage).

Then the amount of data per minute needed across this communication link should be estimated. To estimate the amount of data, multiply the types of conversation that occur across the link by the number of times per hour they are expected to occur. The actual capacity and robustness should be recorded for existing links. It is important to determine carefully each portion of a communication link and list each portion separately. For example, there will be a communications link from a central server to a central network link facility, and from this facility there will be links to other servers. The importance of the first link is very high, as an outage of this link affects all connections from all other points to this central server, possibly causing a massive outage.

When recording robustness, consideration should be given to the following:

- the type of physical medium the link uses, such as fibre optic, microwave or copper wire;
- the provisioning of the link, such as E1, DS3, broadband, ISDN, 56K or cellular data;
- whether the link has redundancy;
- the estimated percentage of availability (uptime);
- the estimated maximum outage time per outage;
- who provides service to repair outages;
- whether any improvements in the link are required to meet the system performance objective; if so, these same methods should be used to estimate the cost of any required link improvement.

These site-to-site pairings with the maximum outage time and anticipated data volume are estimated service levels. It is important that this information is presented to stakeholders for agreement on anticipated costs and the noted maximum outage times. There will almost certainly be times in the future when various systems are unavailable, and acceptable outage parameters should be in agreed in advance. Annual review of performance and acceptable outage parameters is encouraged.

Costs of communications links are highly variable, so creating a table of the estimated costs of the various links available by district is helpful.

Communications system national considerations

System health monitoring and notification

The need to monitor all communications links and distributed systems should be considered, as these can easily number in the hundreds, and it is very difficult otherwise for anyone to resolve problems within an acceptable timeframe. A number of commercial and public domain software packages provide link and server monitoring capabilities and can provide graphs and SMS or email notifications when certain events occur, such as:

- server outage;
- high use of server;
- low free space on server disk drive;
- link outage;
- high use of link;
- low performance of link.

Typically, these systems provide a map offering a visual overview of the health of the system, and statistical and numerical reporting of, for example, use and performance so that future requirements can be estimated.

Data compression

Hardware and software data compression can be used to improve the ability of the communications link to process transactions. Depending on the type of data, data compression can cut the overall volume by half or more. Typically, lossless compression

yields a 25–35% rate, so a typical communication link with compression can process about 1.4 times as much data as a link with no compression. The costs of compression versus the costs of a larger communications link should be evaluated. The comparison should include:

- costs of the hardware or software needed to achieve the compression;
- costs of maintaining the compression hardware and software;
- any anticipated additional outages that may be introduced by the additional systems;
- costs of a higher-capacity communications link (if available) to accommodate the data needs without compression.

Often, there is excess capacity in links used in larger cities. The simplicity of using a link without the additional complexities of compression systems is a compelling factor. The ready availability of technical personnel to identify outages in a more complex environment, and the longer times to diagnose and repair more complex links, should be considered.

In more remote, resource-constrained locations, communications facilities tend to be less capable and less robust, so the task of engineering connections that are capable of providing the needed throughput and are available as needed can be challenging. If communications facilities capable of providing connectivity to suit a near-real-time model are not available, then a batch mode of queuing up requests and responses should be engineered. The communication capabilities within a country play a significant part in determining software technical requirements.

Transaction processing

As enrolments are entered into the system, enrolment processing and administration, verifications of identity and any other NHID events occur. These events generate data transactions, which must be sent to the central registry database, and appropriate acknowledgement of the transaction is sent back.

At the very minimum, a queue of transactions to be sent, transactions that have been sent, and responses that have been sent back from the central registry server must be maintained on the remote site. Depending on the time delay encountered before receiving the responses at the remote site, the overall system is considered to be typically in the range of:

- near-real-time transaction processing, where responses are typically within a number of seconds;
- small batch mode transaction processing, where requests are sent periodically, typically within 15 minutes, and responses typically occur on a batch basis in return;
- large batch mode transaction processing, which is similar to small batch mode, except the batches are larger and occurs within a certain number of hours; this could be several times a day, daily, every few days or once a week. The purpose of this speed of processing is to allow for

intermittent or slower communication links such as cellular links and sending batches of data via a CD-ROM, USB memory stick or other storage medium.

A transaction is not complete until the response has been received at the remote site and processed. Until this occurs, the person entering a NHID application will not know whether the application is being evaluated centrally, approved, rejected or otherwise. Therefore, entering an application at a remote site, batching in less frequent batches of a week, sending applications on a transportable medium to the processing location, and receiving a batched result once a week could mean that responses take several weeks.

In most locations, a hybrid model of near-real-time processing for centrally located enrolment sites and small or large batch processing is required. It is unusual for a country to have excellent connectivity to every remote enrolment site. Locating the enrolment sites more centrally or exclusively centrally will result in a much lower dependency on adequate country-wide communications, but at the expense of being less convenient for people to enrol and return to pick up their NHID cards. Lack of availability of service, such as local computer communication outages, requires the use of batch processing and requires alternative procedures for tracking and caring for patients who present at clinical facilities.

Optimal transaction sizes

It is important to engage experts knowledgeable in transaction processing so they can estimate data volume from each site, communication links and transaction capabilities to develop a processing model that functions as expected.

One of the considerations for estimating data communication volumes is the amount of data associated with each transaction and response. Use of software standards for health data electronic data exchange, such as Health Insurance Portability and Accountability Act (HIPAA) in the United States or HL7, will increase the data communication volume. The use of data processing standards is encouraged but must be evaluated for reasonableness and feasibility; the size, complexity and efficiency of the transactions in a system such as this must fit the capabilities of the communication systems in the country and must not be so complex that it is not feasible to implement the system.

In general, sending batches of data tends to be more efficient than processing the same number of requests as individual transactions, because each group of one or more transactions requires identification information about the sender and pertinent security envelope information; each reply contains similar identification information and security envelopes, and the responses to the requests. This extra overhead is not important where communications bandwidth is adequate to accommodate whatever is needed, but it could be an important issue for a communications resource-constrained setting.

Consideration must be given to the trade-off of efficiency versus use of international standard protocols. Some protocols, such as HL7, tend to be verbose and therefore require much larger transaction sets and higher-speed communication links than a highly optimized transaction protocol. It is always considered to be best practice to use accepted standards, but it has to be reasonable to do so within the circumstances of the country.

Therefore, the final step in deciding the transaction protocol to be used is to estimate the volume to and from each site by building a table of transactions, estimated volume of each type and estimated size of each transaction (using the chosen protocol) and calculating the minimal site-to-site communication bandwidth requirements for the transactions. Since this is such a critical matter, we encourage the engagement of an experienced value-added network engineer for this portion of the planning.

Authentication key management

The purpose of an authentication key is to:

- identify from where a transaction file has come;
- show the data are from a genuine source and have not been modified in any way.

To manage the issuance, expiration and distribution of keys, a program for key management is used. Costs are typically US$ 20–200 per key. The program comes with the keys when issued by an authentication key vendor such as Verisign.

National personnel requirements

- The national coordinator reports to the ministry of health deputy director or equivalent, is the primary overall project manager, and is in charge of resources and interfacing with the ministry and other stakeholders.
- The public policy analyst reports to the national coordinator and is responsible for analysing national legislation and public policies that affect the use of public health data, for confidentiality and privacy of health data, and for providing gap analysis and alternative analyses for the project.
- The project manager reports to the national coordinator and provides operational management and communication for the project. The project manager should be empowered to make operational decisions and needs excellent organizational and communications skills.
- Business analysts report to the project manager and perform scenario evaluation and other situational analyses to assist in the management decision-making process. There should be one to six business analysts for the project, depending on the size of the country and the aggressiveness of the project schedule.
- Staff are needed to receive NHID applications, distribute NHID cards, and provide quality control for NHID applications. The number of these members of staff depends on the number of application points and anticipated volume per location. On average, one person can receive and check the quality of about 50 applications per hour.
- Staff are required to distribute the applications. Regardless of where NHID applications are received, there must be an orderly procedure to ensure forms are not lost. Forms

should be bundled into a batch, the batch signed for, the batch transported to the processing location, and the batch signed for by an authorized person at the processing location.

- Processing/approval personnel review NHID applications and run a script on the system to cross-check for duplicates.
- Each regional office needs at least one, and each national office dozens, of mid- to senior-level data analysts with significant training in the proper resolution of data conflicts, possible duplication, missing information and so on.
- Server and communication systems engineers report to the project manager, provide technical engineering for the hardware and communications systems, and provide technical troubleshooting for the operational communications systems and servers.
- At least one senior-level data system software engineer is needed to evaluate and design software from use cases, and work with the software development staff to ensure proper coding to achieve the goals.
- The transaction (value-added network) engineer has experience in value-added network and data transactions and provides critical recommendations on communication bandwidth requirements and the core application programming interface and formulation of the transaction files.
- Software development engineers are charged with developing the routines, screens, menus and other software processes and features needed to process data efficiently.
- Subject matter experts are highly knowledgeable in a particular field and can provide needed expertise in a particular effort. These are typically people who have been performing these functions for a number of years.
- Testing personnel may be subject matter experts or experienced computer users with adequate time to perform in-depth testing.
- Database administrators are experienced in installing, configuring and operating SQL databases.

Operational support

When the NHID project nears the maintenance and operational phase, some new national information technology support personnel will be needed, including the following:

- Call centre junior helpdesk staff answer questions that arise as applications are accepted. These people typically perform level 1 resolution.
- The call centre lead has experience in providing helpdesk services and ensures the correct support occurs in a timely fashion.
- The junior information technology engineer is sufficiently skilled to be able to identify system issues that

might arise and engage resources to mitigate the problem.

- The senior information technology engineer has experience in large-scale registration systems and a good knowledge of the national system and the underlying technologies in use to deliver the services. This engineer may also be a database administrator.
- Database administrators have experience operating distributed and large-scale SQL databases and can perform all maintenance and operational duties, including performance monitoring and tuning, data backups, and writing all scripts needed to provide an optimum, highly available system.

Setting up NHID application locations

Locations for accepting NHID applications are typically made operational through the following process:

1. Numbers of locations and counts of locations in each province/region/district/city are determined in order to meet the implementation schedule.
2. Specific locations are determined and prioritized.
3. Any construction requirements are met, and locations are reviewed and certified.
4. NHID applications receiving acceptance criteria are determined.
5. Training materials are written for receiving and processing the applications.
6. Applications acceptance personnel are trained.
7. Locations are announced to the public with announcements, signage or other methods.

Receiving a NHID application

Once an application has been received, it should be evaluated for completeness while the applicant is at the facility. When entered into the software system, the application should be evaluated by the system according to the previously established acceptance criteria to determine whether the application meets the minimum criteria. If the system is implemented to provide duplicate checking at the time of application reception, it will provide a secondary check, which should improve efficiency in processing the applications.

Registry processing and administration

There are five stages to processing and completing a registry application:

1. A batch of one or more new applications is received and entered into the registry of new application forms.
2. A series of reports are run to examine the new applications for completeness and duplication in the system. The

output of this report may be simply a printed result, or it may also include output to a batch list of applications that are determined as approved, to be reviewed (giving the reason), and to be denied because they appear to be duplicates.

3. Applications to be approved are marked as such in the system, or the previous reports may produce a group to be approved.

4. Applications that have been approved are printed, sorted and queued for distribution.

5. Cards are distributed and delivered to the applicants. In a highly distributed model, printing may occur at the site where the person is seen; the card is then filed in an appropriate location and presented to the person the next time they present at the location.

- In the highly centralized model, cards are printed centrally on a large-scale heavy-duty printer and sent back to the point of service to be distributed to patients. Cards can be distributed at the next appointment or on an outreach basis. In this model, the printing costs are lower but the distribution costs are higher. This model requires cards to be delivered to district-level locations and relies on existing service providers such as nurses, doctors and community outreach staff distributing the cards to patients.

- In the decentralized model, cards are printed at district facilities and distributed to patients at their next visit, during outreach sessions or via community workers.

- In the highly decentralized model, cards are printed at the point of patient service and distributed at the next visit or via community workers.

Printing and distributing NHID cards

Typical NHID cards are plastic cards similar to driving licence cards. NHID cards may contain a magnetic strip so that basic identifiers such as name, birth date and sex can be read electronically. The use of magnetic strips increases the printing costs because of the additional costs of encoding machines and extra personnel time to print and encode the cards. In 2011, magnetic encoders cost in the range US$ 1200–3000 and added 30 seconds to the time needed to prepare each card.

There are a number of models for printing and distributing NHID cards:

Printing costs are lower when printing centrally, since fewer large-scale heavy-duty printers are required. The cost of such printers is three to four times that of slower, lighter-duty printers. Printers are rated in expected duty cycles and must have at least the capacity to reliably print the required number of cards or more. It is typically more expensive to have outages and repair bills on a printer than to buy a printer with the suitable duty cycle.

Distribution costs are aligned with the number of people involved. If cards are printed centrally, they need to be sent or shipped back to at least the district level. If card distribution is via community workers

who visit patients' residences routinely, there is a small incremental cost of distributing the cards; these costs are mainly time, estimated at 10 minutes per card to organize the card, get it to the community worker visiting the patient, and present the card to the patient with a brief explanation. If visits to patients are primarily to hand out NHID cards, then the costs are higher, as there will be additional time and fuel costs.

To calculate the time needed to distribute cards at the district level, first work out how many patients a day the outreach worker visits. Multiplying that number of patients by 10 minutes indicates the amount of time involved. If additional costs are involved with community workers visiting patients, they can be calculated similarly.

System architecture

Overall system design, including identifying necessary communications links and setting parameters for those links, is the responsibility of the systems architect. The systems architect works with the value-added network engineer to understand the characteristics of the transactions. The systems architect also works with the appropriate staff to provide projections of enrolments, enrolment administration and resolution, and system requirements of the centralized, semi-distributed and fully distributed deployment models.

Equipment requirements

Data entry workstations

Each person putting the application forms into a computer needs a data entry terminal unless scannable forms are used. If scannable forms are used, each site needs at least one workstation paired with a forms scanner.

Biometric workstations

Each site receiving applications needs at least one biometric workstation. This workstation is where photographs are taken, fingerprints read and so on to aid the subsequent identification of patients. Since this can be a time-intensive process, consideration should be given to having a workstation for each person receiving applications so that long queues do not occur.

Central registry servers

At least one larger-capacity server located centrally is needed to host the NHID patient registry. In addition, at least one larger-capacity server located centrally is needed to receive and process transaction files and generate response files.

Distributed registry servers

Most countries will not want to rely on having high-bandwidth communications links and a single server being operational at all times and having the capacity to process all requests of the NHID registry. This is a significant demand and usually means that failure of a single item such as a communications link or router causes widespread outage. One way to avoid the problem of a single point of failure is to use regional servers; an outage then affects only one region and entails a simpler situation for

troubleshooting and repair. Additional costs are related to the additional servers and the possible additional bandwidth requirements of each communications link from each of the additional servers to the central server being significantly higher than they would be otherwise. The reason for the large jump in bandwidth requirements is that everything that is updated on the central database has to be replicated on each regional database; for example, if a country has five regional servers, then each of the five links needs the same high bandwidth as the national server. Typically these links need to be at least 1 megabit but may be up to 10 megabits in larger countries.

Software development

The following is a basic list of software development best practices. This document does not attempt to cover this subject in depth, and topics are discussed only to help estimate the costs and requirements of the project. The required costs and resources cannot be estimated in advance of performing Steps 1–3 or at least a high-level version of Steps 1–3. More accurate estimates can be performed once a project plan has been assembled by the project manager. The basic steps are as follows:

1. Gathering and agreeing on requirements is essential to a successful project. Requirements are broken up into two kinds: functional and non-functional. A good way to document functional requirements is to use use-case analysis.

2. Choosing the appropriate architecture for the application is essential, since the system will provide a critical new resource nationwide. Architectural components include all of the items needed to provide the service, such as servers, workstations, printers, software and communications facilities.

3. Design of the application should include menus, screens, behaviour and outputs. This step should include subject matter expert review of the forms, reports and functionalities of the system. This will constitute a set of system specifications that will ease the job of the code developers.

4. From Steps 1–3, a project plan with estimated resources, costs and calendar can be developed. In addition, the project charter and communications plan should be developed. From this project plan, all resources should be managed and communications regarding progress developed.

5. Code development includes the development of software products specified in the requirements phase. Best practice for constructing code includes the daily or weekly build and unit test.

6. It is important to peer review other people's work. Experience has shown that problems are eliminated earlier this way and reviews are as effective as or even more effective than testing. Any artefact from the development process is reviewed, including plans, requirements, architecture, design, code and test cases.

7. Testing is an integral part of software development that needs to be planned. It is important that testing is done proactively, meaning that test cases are planned before coding starts and test cases developed while the application is being designed and coded. Subject matter experts should be involved in testing all forms and reports.

8. A method to catch some architectural defects is to simulate load testing on the application before it is deployed and to deal with performance issues before they become problems.

9. It is important to establish quality priorities and release criteria for the project so that a plan is constructed to help the team achieve good-quality software. As the project is coded and tested, the defect discovery and fix rate can help measure the maturity of the code. All of these items should be tracked in a defect-tracking database.

10. Once the code has been tested and a production date set, the production system should be set up and all appropriate data migrated into the production system. Final testing of the forms and reports should be performed by the subject matter experts and quality-testing personnel.

11. Once the production system has been installed and tested, a deployment date should be set. All technical personnel should be available for the initial production days in order to quickly resolve any unforeseen problems.

12. Once the system is in production, ongoing support personnel will be required, such as systems administrators, database administrators and communications engineers.

There are many other good sources for information on project management, including:

- http://en.wikipedia.org/wiki/Project_management
- http://www.pmi.org/PMBOK-Guide-and-Standards.aspx
- http://www.projectmanagement.com

Training requirements

Performing initial training and ongoing new employee and refresher training can be a significant undertaking, both financially and logistically. Regardless of the manner in which the training is performed, there should be comprehensive training for:

- application review and acceptance;
- data entry of application;
- registry administration, including first-level reporting, maintenance and anomaly review;
- registry second- and third-tier resolution;
- NHID issuance and distribution.

Training on location and at central locations

Training on location is the most efficient way from the site perspective to train employees. If the site has a location at which the training can occur, then a computer-based training course may be the most cost-effective way of providing training. This course may be developed by, for example, video- and audio-recording a trainer. If computer-based training is used, a chapter-by-chapter test should be part of the training video and the testing results stored in a way that ensures the integrity of the testing process.

Training at a centralized location requires larger training facilities but can be the least expensive way of proving training, if the travel and per diem costs involved in bringing the personnel to the location are low; often, however, the travel and per diem costs for all the people to be trained far exceed the costs of sending a trainer to a region or district and doing the training there.

Regardless of the method chosen, there will be ongoing training requirements due to scale-up and personnel turnover. Consideration should be given as to how to provide these services over time.

Costing

Estimating the financial requirements for a NHID project is a complex task with highly variable resource costs. The items to cost include:

- office space, computer equipment, staffing, staff training and communications lines from each health identification application point to the central location;
- office space, computer equipment, staffing and staff training for each central location processing the applications;
- office space, computer equipment, communications links, staffing and staff training for each location manufacturing NHID cards;
- office space, computer equipment, software, communications links, staffing and staff training for the systems programming group;
- office space, computer equipment, software, communications links, staffing and staff training for the systems support group;
- computer hardware, software and communications links for all points where servers will be located;
- a senior project manager and any consultants to develop, manage and communicate all project plans;
- a technical working group to evaluate the status of the project, and a management committee to provide oversight and management decisions for the project.

Summary

Implementation of a NHID system is a complex, time-consuming task. For the project to have a successful outcome, a reasonable amount of resources must be available. The easiest way to lower costs is to implement the simplest model over a generation of time; that is, as people reach a certain age, they are issued a NHID card and added into the national registry. The most complex model is an aggressively implemented, highly distributed model.

Annex 6
Annotated bibliography

This annotated bibliography is a 2012 updated edition of the bibliography prepared as background documentation to inform the participants of the 2009 workshop on unique identifiers held in Montreux, Switzerland.

Introduction

Purpose of the bibliography

An effective response to the HIV epidemic is longitudinal in nature and multi-sectoral 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 people living with HIV 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 of UNAIDS and PEPFAR convened a workshop in 2006 that produced the *Interim guidelines on protecting the confidentiality and security of HIV information*.[3]

Collecting accurate patient data is complicated by the multitude of data sources, disparity in technology, inconsistencies in data storage formats between facilities and over time, shifting data needs, and absence of unique patient identifiers. Building on the 2006 workshop, UNAIDS and PEPFAR convened the Health Services Identifier Workshop in Montreux, Switzerland, 24–26 February 2009. This workshop brought together 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 bibliography was originally developed to provide a review of existing work on NHIDs and present experiences with such identifiers from several countries. The bibliography was intended to be used as a resource for the workshop, presenting material that workshop participants could review and critique and use to inform the development of guidelines for the design and implementation of NHIDs in middle- and lower-income countries. The bibliography was updated in 2012 to include relevant recent publications.

Approach to the bibliography

Macro International Inc. was contracted to search the literature and put together the 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 published literature and industry information found through websites. 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 2009 workshop and included unique health identifiers, HIV longitudinal monitoring, privacy, disease

3 Interim guidelines on protecting the confidentiality and security of HIV information: proceedings from a workshop, 15–17 May 2006, Geneva, Switzerland. Geneva: Joint United Nations Programme on HIV/AIDS and United States President's Emergency Plan for AIDS Relief; 2007 (http://data.unaids.org/pub/manual/2007/confidentiality_security_interim_guidelines_15may2007_en.pdf).

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.

Located articles were distributed to each member of the review team. Each article or document was reviewed, selected for inclusion or exclusion, classified independently, and 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.

Content and format

The bibliography is divided into two sections – recommended reading and further reading. The list of 18 articles for recommended reading includes articles from all 5 categories (unique identifiers, record linkage, technology, privacy and implementation). The 18 articles are recommended because they include concise overviews of the key topics with a sufficient level of detail.

The articles included as further reading provide additional context and detail and differing viewpoints.

The references are listed by author, in alphabetical order. Each reference includes the full reference details, an abstract and a URL to locate the article online. If an article is missing an abstract, the reviewer has summarized the article and included a brief synopsis in lieu of an abstract.

Access to the articles

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

Notes to the 2012 revised edition

This revised edition of the bibliography contains additional sources of information that have become available since the workshop. New entries are indicated with an asterisk (*). Some of the key new entries include:

- a relevant ISO standard;
- in the privacy and confidentiality section, a new peer-reviewed paper and commissioned research report;
- in the biometrics section, a peer-reviewed paper on data fusion of combined or staged biometric data.

Small grammatical and typographical corrections have been made in some existing items.

General recommended reading

Standard guide for properties of a universal healthcare identifier (UHID). West Conshohocken, PA: American Society for Testing and Materials; 2007 (ASTM E-1714-00; http://www.astm.org/Standards/E1714.htm).

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

Guide for implementation of a voluntary universal healthcare identification system. West Conshohocken, PA: American Society for Testing and Materials; 2007 (ASTM E2553; http://www.astm.org/Standards/E2553.htm).

> This document describes the implementation principles needed to create a voluntary universal health-care identification system. The purpose of this system is to enable unambiguous identification of individuals in order to facilitate the delivery of health care. The voluntary universal health-care identification system should be dedicated exclusively to the needs and functions of health care. The system is designed to represent no or minimal increased risk to health-care 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 health care. 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, if any, of the safety concerns associated with its use.

*Beck EJ, Mandalia S, Harling G, Santas X, Mosure D, Delay P. Protecting HIV information in countries scaling up HIV services: a baseline study. J Int AIDS Soc. 2011;14:6 (http://www.jiasociety.org/index.php/jias/article/view/17444).

> **Background:** individual-level data are needed to optimize clinical care and monitor and evaluate HIV services. Confidentiality and security of such data must be safeguarded to avoid stigmatization and discrimination of people living with HIV. The study's goal was to assess the extent that countries scaling up HIV services have developed and implemented guidelines to protect the confidentiality and security of HIV information.
>
> **Methods:** questionnaires were sent to UNAIDS field staff in 98 middle- and lower-income countries, some reportedly with guidelines (G-countries) and others intending to develop them (NG-countries). Responses were scored, aggregated and weighted to produce standard scores for six categories: information governance, country policies,

data collection, data storage, data transfer and data access. Responses were analysed using regression analyses for associations with national HIV prevalence, gross national income per capita, Organisation for Economic Co-operation and Development (OECD) income, receipt of PEPFAR funding, and being a G- or NG-country. Differences between G- and NG-countries were investigated using nonparametric methods.

Results: higher information governance scores were observed for G-countries compared with NG-countries; no differences were observed between country policies or data-collection categories. For data storage, data transfer and data access, G-countries had lower scores compared with NG-countries. No significant associations were observed between country score and HIV prevalence, per capita gross national income, OECD economic category and whether countries had received PEPFAR funding.

Conclusions: few countries, including G-countries, had developed comprehensive guidelines on protecting the confidentiality and security of HIV information. Countries must develop their own guidelines, using established frameworks to guide their efforts, and may require assistance in adapting, adopting and implementing these guidelines.

Churches T, Christen P. Some methods for blindfolded record linkage. BMC Med Inform Decis Mak. 2004;4:9 (http://www.biomedcentral.com/1472-6947/4/9).

Background: the linkage of records that 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 that 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 that permits record linkage studies to be carried out with negligible risk of invasion of personal privacy.

*Hammond E, Bailey C, Boucher P, Spohr M, Whitaker P. Connecting information to improve health. Health Affairs 2010;29:284–8 (http://content.healthaffairs.org/content/29/2/284.full.pdf).

> Effective health information systems require timely access to all health data from all sources, including sites of direct care. In most parts of the world today, these data most likely come from many different and unconnected systems but must be organized into a composite whole. The word interoperability captures what is required to accomplish this goal. The five priority areas for achieving interoperability in health-care applications are patient identifier, semantic interoperability, data interchange standards, core datasets and data quality. Differences are contrasted in developing and developed countries. Important next steps for health policy-makers are to define a vision, develop a strategy, identify leadership, assign responsibilities and harness resources.

Hieb BR. The case for a voluntary national healthcare identifier. J ASTM Int. 2006;3 (http://www.astm.org/DIGITAL_LIBRARY/JOURNALS/JAI/PAGES/JAI13891.htm).

> This paper describes a proposed system to create a voluntary national health-care identification system for the United States. The 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 system is able to meet the vast majority of the objections that have previously been raised concerning a national health-care 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 before implementation.

Hillestad R, Bigelow JH, Chaudhry B, Dreyer P, Greenberg MD, Meili RC, et al. Identity crisis: an examination of the costs and benefits of a unique patient identifier for the U.S. health care system. Santa Monica, CA: RAND Corporation; 2008 (http://www.rand.org/pubs/monographs/MG753.html).

> Correctly linking patients to their health data is a vital step in creating good-quality health care. The two primary approaches to this linking are the unique patient identifier and statistical matching based on multiple personal attributes, such as name, address and social security number. Lacking a unique patient identifier, most of the United States 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 towards a unique patient identifier 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.

*Health informatics: identification of subjects of health care. Geneva: International Organization for Standardization; 2011 (ISO/TS 22220:2011; http://www.iso.org/iso/catalogue_detail.htm?csnumber=40782).

> This technical specification identifies the data elements and relevant structure and content of the data used to manually identify individuals in a health-care setting and provides support to the identification of individuals in a consistent manner between systems that will support the natural changes in usage and application of the various names used by people over time. This document addresses the business requirements of identification and the data needed to improve the confidence of health service providers and subjects of care identification. It defines the data used to identify subjects of care and the business processes associated with this activity, whether computerized or manual. This document is intended to be used to support both identification of subjects of care by individuals and computerized identification in automated matching systems.

McMahon DJ. The future of privacy in a unified national health information infrastructure. Seton Hall Law Rev. 2008;38:787 (http://www.ncbi.nlm.nih.gov/pubmed/18623904).

> This comment begins with an overview of the current state of health-care privacy law and the need for adequate privacy protection. Part III then describes and analyses selected bills that 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 that uses a new property right in personal information along with contractual and tort-based protection.

Meray N, Reitsma JB, Ravelli AC, Bonsel GJ. Probabilistic record linkage is a valid and transparent tool to combine databases without a patient identification number. J Clin Epidemiol. 2007;60:883–91 (http://www.jclinepi.com/article/S0895-4356(06)00500-2/pdf).

> **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 paediatricians/neonatologists). These registries do not share a 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.

Meux E. Encrypting personal identifiers. Health Serv Res. 1994;29:247–56 (http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=1070001&blobtype=pdf).

Study setting: a state-wide patient discharge database contained only one unique identifier: the social security number. A method was developed to transform (encrypt) the social security number so that it could be made publicly available, for purposes of linking discharge records, without revealing the social security number itself. The method of encrypting the social security number into a record linkage number is described.

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

Netter W. Curing the unique health identifier: reconciliation of new technology and privacy rights. Jurimetrics 2003;43:165–86 (http://medscapecrm.org/medline/abstract/15156882).

The Health Insurance and Portability and Accountability Act 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 multi-tiered system that would be administered by a mixed government and private entity. Consumers could voluntarily opt in to the system.

Nationwide privacy and security framework for electronic exchange of individually identifiable health information. Rockville, MD: Office of the National Coordinator for Health Information Technology, United States Department of Health and Human Services; 2008 (http://www.hhs.gov/healthit/documents/NationwidePS_Framework.pdf).

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 people, 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.

The unique records portfolio: a guide to resolving duplicate records in health information systems. Decatur, GA: Public Health Informatics Institute; 2006 (http://www.phii.org/resources/view/4380/The%20Unique%20Records%20Portfolio%3A%20A%20guide%20to%20resolving%20duplicate%20records%20in%20health%20information%20systems).

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 de-duplication strategies. The concepts, example and tools in the unique records portfolio are intended to help public health agencies and health-care organizations address the de-duplication challenge. The portfolio also articulates the challenges and solutions to developing de-duplication strategies and describes the implications that various approaches have on data use. Case examples of the de-duplication 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.

Sauleau EA, Paumier JP, Buemi A. Medical record linkage in health information systems by approximate string matching and clustering. BMC Med Inform Decis Mak. 2005;5:32 (http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1274322).

Background: multiplication of data sources within heterogeneous health-care information systems always results in redundant information, split among multiple databases. The 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 standardize 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 3000 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.

Stewart SP, Arellano MB, Simborg DW. Optimal patient identification system. J Am Med Rec Assoc. 1984;55:23–7 (http://www.ncbi.nlm.nih.gov/pubmed/10310711).

Health facilities continually seek ways to provide services more efficiently. Linking all facets of patient care through a well-designed patient identification system can be expected to increase the efficiency of patient care. Linking patient service data to financial data allows administrators to analyse 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.

Tierney WM, Beck EJ, Gardner RM, Musick B, Shields M, Shiyonga NM, et al. Viewpoint: a pragmatic approach to constructing a minimum data set for care of patients with HIV in developing countries. J Am Med Inform Assoc. 2006;13:253–60 (http://www.jamia.org/cgi/content/abstract/13/3/253).

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

Further reading: unique identifiers

Appavu SI. Analysis of unique patient identifier options final report. Washington, DC: United States Department of Health and

Human Services; 1997 (http://ncvhs.hhs.gov/app0.htm).

The Health Insurance Portability and Accountability Act of 1996 requires the Secretary of the United States Department of Health and Human Services 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 Department of Health and Human Services by the National Committee on Vital and Health Statistics.

Appavu SI. Unique patient identifiers: what are the options? J AHIMA. 1999;70:50–57 (http://www.ncbi.nlm.nih.gov/pubmed/10977406).

In 1997, the United States Department of Health and Human Services commissioned a study to analyse the various patient identification systems available. The study consisted of an objective analysis of the various unique patient identifier options available for use in the health-care 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.

Australia Standard Committee IT-014 Health Informatics. Health care client identification. Sydney: Standards Australia; 2006 (AS 5017-2006; http://www.saiglobal.com/PDFTemp/Previews/OSH/as/as5000/5000/5017-2006.pdf).

The objective of this standard is to provide the health industry with a specific standard for health-care client identification for clinical and administrative data management purposes (data structure and specification) that promotes uniformly good practice in identifying individuals and recording identifying data so as to ensure that each individual's health record is associated only with that individual. The standard also provides the basis for future linkage of data as authorized by law and appropriate for clinical management of patients and statistical research purposes.

Fernandes L, O'Connor M. Patient identification in three acts. J AHIMA. 2008;79:46–9 (http://library.ahima.org/xpedio/groups/public/documents/ahima/bok1_037463.hcsp).

This article discusses three solutions for patient identification that appear to be emerging ahead of the pack: probabilistic matching, national identifiers and voluntary universal health identifiers. The article notes recent developments, such as the United States presidential election and the release of the RAND report, which may have an impact on patient identification standards.

France FH, De Clercq E, Bangels M. Purposes of health identification cards in Belgium. In: Engelbrecht R, Geissbuhler A,

Lovis C, editors. Connecting medical informatics and bio-informatics. Amsterdam: IOS Press; 2005;116: 415-420 (http://iospress.metapress.com/content/0v63l16jdx50py30).

Although other alternatives may 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, and for health information exchange and database querying. Local applications may exist independently, but the federal state has now developed Be-Health, a platform for health professionals and social security personnel and the public to facilitate a common access to some health data. Security conditions have been defined and are described.

Gabrieli ER. Guide for unique healthcare identifier model. J Clin Comput. 1993;21:101–39 (http://www.ncbi.nlm.nih.gov/pubmed/10126775).

This guide presents the current understanding of the role and construction of a unique health-care identifier intended to be limited strictly to use within the American health-care system. The guide sets forth the fundamental criteria for a unique health-care identifier that can facilitate linking computer-based records on the same patient, provides for data security measures for confidential clinical data, and identifies patients receiving clinical care.

Quantin C, Allaert FA, Avillach P, Fassa M, Riandey B, Trouessin G, et al. Building application-related patient identifiers: what solution for a European country? Int J Telemed Appl. 2008 (http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2288643/).

The authors propose a method using a derived social security number with the same reliability as the social security number. They show the anonymity techniques classically based on unidirectional hash functions such as the secure hash algorithm (SHA-2) function, which 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 the 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.

Rollins G. This year's models: a look at patient ID in the four newly demonstrated NHIN prototypes. J AHIMA. 2007;78:34–7 (http://library.ahima.org/xpedio/groups/public/documents/ahima/bok1_033608.hcsp).

This winter the health-care industry got a first look at four prototypes for a nationwide health-care data exchange network. Created under contract to the federal government, the demonstration projects paired technology developers

and health-care 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.

Further reading: record linkage

Alemi F, Loaiza F, Vang J. Probabilistic master lists: integration of patient records from different databases when unique patient identifier is missing. Health Care Manag Sci. 2007;10:95–104 (http://www.springerlink.com/content/c0n30483l32421r6).

We show how Bayesian probability models can be used to integrate two databases, one of which does not have a key for uniquely identifying patients, such as social security number or medical record number. The analyst selects a set of imperfect identifiers, such as diagnosis at last visit and first name. 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 patient from the likelihood ratios.

We test that the procedure is effective by examining data from the Medical Expenditure Panel Survey Population Characteristics dataset, a publicly available dataset. We randomly selected 1000 cases for training dataset – these constituted the known cases. The algorithm was used to identify whether 100 cases not in the training dataset 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 classified correctly 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 classified correctly.

These data suggest the accuracy of our automated and mathematical procedure to merge data from two different datasets 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.

Brum L, Kupek E. Record linkage and capture-recapture estimates for underreporting of human leptospirosis in a Brazilian health district. Braz J Infect Dis. 2005;9:515–20 (http://www.scielo.br/pdf/bjid/v9n6/a11v09n6.pdf).

Record linkage and capture–recapture models were used to estimate the number of cases of human leptospirosis in the health district of Santa Maria in southern Brazil. Twelve months of laboratory, hospital and epidemiological surveillance data were matched by name, age, residence and 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 cases per 1000 inhabitants and a case fatality of 0.37%. Severe cases were found predominantly 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.

Christen P. Febrl: a freely available record linkage system with a graphical user interface. Sydney: Australian Computer Society; 2008 (http://crpit.com/confpapers/CRPITV80Christen.pdf).

Record or data linkage is an important enabling technology in the health sector, as linked data are 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, originating mostly 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 so-called black box commercial software. This makes it difficult for users to learn about new record linkage techniques and to compare existing and new linkage techniques. 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 standardization, 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, and for practitioners to conduct linkages with datasets containing up to several hundred thousand records.

Colias M. Disease registries. Hosp Health Netw. 2005;79:62–8 (http://www.ncbi.nlm.nih.gov/pubmed/15770911).

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

Dunn HL. Record linkage. Am J Publ Health 1946;36:1412–16 (http://ajph.aphapublications.org/doi/pdf/10.2105/AJPH.36.12.1412).

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 organization.

Fellegi IP, Sunter SB. A theory of record linkage. J Am Stat Assoc. 1969;64:1183–210 (http://www.jstor.org/pss/2286061).

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 that represent identical people, objects or events (said to be matched).

A comparison is 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 that make it a practical working tool are given.

Gill L. Methods for automatic record matching and linking and their use in national statistics. Newport: Office for National Statistics; 2001 (http://www.ons.gov.uk/ons/guide-method/method-quality/specific/gss-methodology-series/gss-methodology-series--25--methods-for-automatic-record-matching-and-linkage-and-their-use-in-national-statistics.pdf).

This report concentrates on the use of record linkage for statistical purposes only, to produce summaries and statistics. Many of the administrative and survey datasets are collected under a 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 nontechnical 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.

Gomatam S, Carter R, Ariet M, Mitchell G. An empirical comparison of record linkage procedures. Stat Med. 2002;21:1485–96 (http://onlinelibrary.wiley.com/doi/10.1002/sim.1147/abstract).

We consider the problem of record linkage in the situation where we have only non-unique identifiers, such as names, sex and race, 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, differences, strengths and weaknesses is presented.

Grannis SJ, Overhage JM, McDonald CJ. Analysis of identifier performance using a deterministic linkage algorithm. Proc AMIA Symp. 2002;305–9 (http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=2244404&blobtype=pdf).

As part of developing a record linkage algorithm using de-identified patient data, we analysed 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 analysed samples from each registry totalling 6000 record-pairs to establish a linkage gold standard. Using the 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 sex. This found 87% and 88% of the links without any false links. We achieved sensitivities of 90–92% while maintaining 100% specificity using combinations of social security number, sex, 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.

Grannis SJ, Overhage JM, Hui S, McDonald CJ. Analysis of a probabilistic record linkage technique without human review. Proc AMIA Symp. 2003;259–63 (http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=1479910&blobtype=pdf).

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 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 as 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 expectation maximization 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.

Kelman C. The Australian National Death Index: an assessment of accuracy. Aust N Z J Publ Health 2000;24:201–3 (http://www3.interscience.wiley.com/journal/119012667/issue).

> **Objective:** the Australian National Death Index 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 National Death Index 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 = 2990). Indices of accuracy for the National Death Index were calculated.
>
> **Results:** for this particular study, the National Death Index 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 people using the National Death Index 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.

Li B, Quan H, Fong A, Lu M. Assessing record linkage between health care and vital statistics databases using deterministic methods. BMC Health Serv Res. 2006;6:48 (http://www.biomedcentral.com/content/pdf/1472-6963-6-48.pdf).

> **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, given name, sex and date of birth) were used to determine the optimal combination. The correct linkage rate was assessed using a 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.

Lui S, Wen SW. Development of record linkage of hospital discharge data for the study of neonatal readmission. Chron Dis Inj Can. 1999;20:77–81 (http://www.phac-aspc.gc.ca/publicat/cdic-mcc/20-2/c_e.html).

> Computerized record linkage has been used increasingly in epidemiological studies. We developed a multistage, deterministic matching algorithm using various combinations of key variables.

Then, from the records for 1 March 1993 to 31 March 1996, contained in the discharge abstract database of the Canadian Institute for Health Information, 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 (steps 2 and 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 epidemiological studies of other short-term events.

Lynch BT, Arends WL. Selection of a surname coding procedure for the SRS record linkage system. Washington, DC: United States Department of Agriculture, Sample Survey Research Branch, Research Division; 1977 (http://www.nass.usda.gov/research/reports/Internet_Yield/reportsxyield.html).

The Statistical Reporting Service 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 will be selected from this list. This system uses a probability model that 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, the Statistical Reporting Service 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 pair-wise 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 Statistical Reporting Service System and thereby reduce the number of comparisons.

Machado CJ. A literature review of record linkage procedures focusing on infant health outcomes. Cad Saude Publica 2004;20:362–71 (http://www.scielosp.org/pdf/csp/v20n2/03.pdf).

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 datasets and provide a critical review of the studies in Brazil.

Machado CJ, Hill K. Probabilistic record linkage and an automated procedure to minimize the undecided-matched pair problem. Cad Saude Publica 2004;20:915–25 (http://www.scielo.br/pdf/csp/v20n4/05.pdf).

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 Sao 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 2249 to 2827. 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.

Markle Foundation, Robert Wood Johnson Foundation. Linking healthcare information: proposed methods for improving care and protecting privacy. New York: Markle Foundation; 2005 (http://www.markle.org/publications/863-linking-health-care-information-proposed-methods-improving-care-and-protecting-priv).

The linking of vital information as patients receive care from a fragmented health-care system is a problem that has consistently plagued interoperability efforts in health care. 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.

Miller PL, Frawley SJ, Sayward FG. IMM/Scrub: a domain-specific tool for the deduplication of vaccination history records in childhood immunization registries. Comput Biomed Res. 2000;33:126–43 (http://www.sciencedirect.com/science/journal/00104809).

IMM/Scrub is a pilot tool developed to assist in the de-duplication of vaccination history records in childhood immunization registries. This problem is complicated by a number of factors including that fact that some doses are numbered and some are not, doses may have different dose numbers, doses may specify different preparations within a vaccine series, one dose may indicate a combination vaccine and the other dose may specify one component of that combination, two doses may have slightly different dates, and combinations of any of these problems may occur together. IMM/Scrub is designed to help detect 10 different types of vaccination dose duplicate 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 programme, which can provide additional assistance in the data cleaning process. The paper describes the design of the current pilot implementation of IMM/Scrub, the lessons learned during its implementation, and our preliminary experience applying it to data from three immunization databases, from a state, a metropolitan area and an academic medical centre.

Newcombe HB, Kennedy JM, Axford SJ, James AP. Automatic linkage of vital records. Science 1959;130:954–9 (http://www.sciencemag.org/cgi/content/citation/130/3381/954).

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 that occurred in the Canadian province of British Columbia during the year 1955 (34 138 births) and of the marriages that took place in the same province over the 10-year period 1946–1955 (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.

Newman TB, Brown AN. Use of commercial record linkage software and vital statistics to identify patient deaths. J AMIA 1997;4:233–7 (http://www.jamia.org/cgi/content/abstract/4/3/233).

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 United States Social Security Administration (N = 13 341 581). We linked 96.5% of 3448 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 6444 patients 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 linked accurately with state and national vital statistics using commercial record linkage software.

Pates RD, Scully KW, Einbinder JS, Merkel RL, Stukenborg GJ, Spraggins TA, et al. Adding value to clinical data by linkage to a public death registry. Stud Health Technol Inform. 2001;84:1384–8 (http://iospress.metapress.com/content/lwxvjh0aw623n8l4).

> We describe the methodology and impact of merging detailed state-wide mortality data into the master patient index tables of the clinical data repository 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 clinical data repository to those in the state-wide files using the following criteria: social security number, patient's last name and birth date, and patient's last name and given 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 8-year period as an internal control. We conclude that we are able to update our clinical data repository 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 clinical data repository 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.

Porter EH, Winkler WE. Approximate string comparison and its effect on an advanced record linkage system. In: Record linkage techniques – 1997: proceedings of an international workshop and exposition. Washington, DC: National Academy Press; 1999:190–202 (http://www.census.gov/srd/papers/pdf/rr97-2.pdf).

> Record linkage, sometimes referred to as information retrieval, 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% of given names and in surnames, effective methods for dealing with typographical errors 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 estimated automatically by an expectation-maximization algorithm for latent class, log-linear models of the type arising in the Fellegi–Sunter model of record linkage. Overall matching efficacy is improved further by linear assignment algorithm that forces one-to-one matching.

Potosky AL, Riley GF, Lubitz JD, Mentnech RM, Kessler LG. Potential for cancer related health services research using a linked Medicare–tumor registry database. Med Care 1993;31:732–48 (http://www.jstor.org/pss/3767064 and http://www.jstor.org/pss/3765984).

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 outcomes of people with cancer . 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 nine population-based tumour registries that collect standardized clinical information on cases diagnosed in separate, geographically defined areas covering approximately 10% of the United States population. Using a deterministic matching algorithm, the records of 94% of SEER registry cases diagnosed at age 65 years or older between 1973 and 1989, or more than 610 000 people, were successfully linked with Medicare claims files. The resulting database, combining clinical characteristics with information on use and costs, will permit the investigation of the contribution of various patient and health-care setting factors to treatment patterns, costs and medical outcomes.

Rabeneck L, Menke T, Simberkoff MS, Hartigan PM, Dickinson GM, Jensen PC, et al. Using the national registry of HIV-infected veterans in research: lessons for the development of disease registries. J Clin Epidemiol. 2001;54:1195–203 (http://www.ncbi.nlm.nih.gov/pubmed/11750188).

The purpose of this article is to describe the structure, function, applications and limitations of the United States Department of Veterans Affairs (VA) HIV registry, and to discuss how investigators developing disease registries in the future could benefit from our experience. We examined the number of people with AIDS 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 the VA 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.

Sideli RV, Friedman C. Validating patient names in an integrated clinical information system. Proc Annu Symp Comput Appl Med Care 1991;588–92 (http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=2247599&blobtype=pdf).

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 10 000 patient name pairs. In addition, the types of spelling and typographical error that occur in patient names in the pathology database at California Pacific Medical Center are described. The data analysis shows that the LCS method performs better than the other techniques when compared with manual analysis.

Van den Brandt PA, Schouten LJ, Goldbohm RA, Dorant E, Hunen PM. Development of a

record linkage protocol for use in the Dutch Cancer Registry for epidemiological research. Int J Epidemiol. 1990;19:553–8 (http://ije.oxfordjournals.org/cgi/content/abstract/19/3/553).

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 sex, 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.

Whalen D, Pepitone A, Graver L, Busch JD. Linking client records from substance abuse, mental health and Medicaid state agencies. Rockville, MD: Center for Substance Abuse Treatment and Center for Mental Health Services, Substance Abuse and Mental Health Services Administration (SAMHSA); 2000 (SAMHSA publication no. SMA-01-3500; http://store.samhsa.gov/product/Linking-Client-Records-from-Substance-Abuse-Mental-Health-and-Medicaid-State-Agencies/BKD393).

This report describes the concepts behind record linking and the specific application of record linking in building databases integrating information about mental health and alcohol/drug 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, but although 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 mental health and alcohol/drug 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 website. Potential applications and extensions of this methodology are discussed and future directions are outlined.

Williams BC, Demitrack LB, Fries BE. The accuracy of the national death index when personal identifiers other than social security number are used. Am J Publ Health 1992;82:1145–7 (http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=1695740&blobtype=pdf).

The study analysed 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 people and 92–99% of living people. These results should prove useful in ascertaining the mortality status of patient populations without social security numbers.

Winkler WE. Matching and record linkage. In: Cox BG, et al., editors. Business survey methods. New York: J. Wiley & Sons; 1995: 355–84 (http://www.census.gov/srd/papers/pdf/rr93-8.pdf).

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.

Further reading: technology

Beckwith BA, Mahaadevan R, Balis UJ, Kuo F. Development and evaluation of an open source software tool for deidentification of pathology reports. BMC Med Inform Decis Mak. 2006;6:12 (http://www.biomedcentral.com/content/pdf/1472-6947-6-12.pdf).

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 de-identification tool tailored for pathology reports. We designed a three-step process for removing potential identifiers. Each pathology report was reviewed manually before and after de-identification to catalogue 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 de-identification program that performs well on free-text pathology reports.

Dumortier J. The European Directive 1999/93/EC on a community framework for electronic signatures. In: Lodder AR, Kaspersen HW, editors. eDIrectives: guide to European Union law on e-commerce – commentary on the directives on distance selling, electronic signatures, electronic commerce. London: Kluwer Law International; 1999: 33–65 http://www.amazon.com/Edirectives-European-E-Commerce-Electronic-Commerce/dp/9041117520

This 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.

Fulcher J. The use of patient biometrics in accessing electronic health records. Int J Healthc Technol Manag. 2004;6:20–31 (http://www.inderscience.com/search/index.php?action=record&rec_id=4822&prevQuery=&ps=10&m=or).

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.

Halamka J. Early experiences with positive patient identification. J Healthc Inf Manag. 2006;20:25–7 (http://www.researchgate.net/publication/7342955_Early_experiences_with_positive_patient_identification).

To accomplish the goal of positive identification of patients, staff and medications, Beth Israel Deaconess Medical Center investigated two major kinds of technology: barcodes and radio frequency identification. The article describes different use cases and the pros and cons of each technology.

Patient identifier cross-reference HL7 V3 (PIXV3) and patient demographic query HL7 V3 (PDQV3). Oak Brook, IL: IHE; 2007 (http://www.ihe.net/Technical_Framework/upload/IHE_ITI_TF_Supplement_PIX_PDQ_HL7v3_TI_2007_08_15.pdf and http://www.ihe.net/Technical_Framework/upload/IHE_ITI_TF_Supplement_PIX_PDQ_-HL7v3_-TI_2008-11-11.pdf).

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 to that for the HL7 v2.5 messages (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 with the previous year are minimal, as the message content only changes the focal class from identified entity to patient.

Kardas G, Tunali ET. Design and implementation of a smart card based healthcare information system. Comput Methods Programs Biomed. 2006;81:66–78 (http://ege.academia.edu/GeylaniKardas/Papers and http://www.globalhivmeinfo.org/HIS/Annotated%20Bibliography/Article%209a.pdf).

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

Leonard DC, Pons A. Realization of a universal patient identifier for electronic medical records through biometric technology. IEEE Trans Inf Technol Biomed. 2009;13:494–500 (http://www.ncbi.nlm.nih.gov/pubmed/19273015).

The advent of universally accessible health-care data benefits all participants, but one of the outstanding problems that must be addressed is how the creation of a standardized nationwide electronic health-care record system in the United States would uniquely identify and match a composite of an individual's recorded health-care information to an identified individual patient out of approximately 300 million people to a one-to-one 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: fingerprint; iris; retina scan; DNA. This would allow a patient to have real-time access to all of their recorded health-care information electronically whenever it is necessary, securely with minimal effort, greater effectiveness and ease.

Mathieson S. Smartcards. Smart move. Health Serv J. 2005;115(Suppl. 5972):12–13 (http://www.ncbi.nlm.nih.gov/pubmed/16171136?dopt=Abstract).

This article provides an overview of the use of smartcards for health care 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.

Salkowitz SM, Clyde S. De-duplication technology and practices for integrated child-health information systems. Decatur, GA: Public Health Informatics Institute; 2003 (https://www.hln.com/assets/pdf/dedupe.pdf).

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 de-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 it 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.

*Complementary smart card guidance for the WEDI Health Identification Card Implementation Guide. Princeton Junction, NJ: Smart Card Alliance; 2011 (TR HCC-11002; http://www.smartcardalliance.org/pages/publications-complementary-smart-card-guidance-for-the-wedi-health-identification-card-implementation-guide).

For organizations considering upgrading their member identity cards to smartcards, this document serves as a supplement to the WEDI Health Identification Card Implementation Guide. It provides WEDI-compliant smartcard designs and includes a discussion of the features and benefits of smart identity cards for health-care providers and payers.

Wilson S. 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, 2005 (http://lockstep.com.au/library/privacy/anon-health-ids).

Default thinking about electronic health records and unique health identifiers has

settled on a national numbering scheme, despite the fact that patient privacy can be seriously jeopardized 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 does anything to control the underlying linkages of identifiers and names, and so a serious gap persists in electronic health records strategy and architecture. This paper presents a new way to fundamentally anonymize unique health identifiers through a novel use of public key certificates and smartcards. The design presented here secretes each unique health identifier within an anonymous digital certificate and links one or more certificates to a smartcard. If an electronic health record entry is digitally signed via such a certificate, then that entry is directly linked to the unique health identifier 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 offline identity resolution, reduced reliance on centralized, mission-critical identity servers, seamless support for multiple electronic health records, and compatibility with a range of smartcard choices available to consumers in the near future.

Further reading: legal, ethical and privacy considerations

Issues for the use of unique patient identifiers in statistical collections. Canberra: Australian Institute of Health and Welfare; 2002 (http://www.aihw.gov.au/publications/hwi/iuupisc02/iuupisc02.pdf).

> 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 dataset management raised by proposals for the introduction of unique patient identifiers. The discussion covers unique patient identifiers with and without explicitly identifying details such as names and addresses.

*Brown CL. Health-care data protection and biometric authentication policies: comparative culture and technology acceptance in China and in the United States. Rev Policy Res. 2012;29 (http://onlinelibrary.wiley.com/

doi/10.1111/j.1541-1338.2011.00546.x/abstract).

> A proliferation of health information technology policies to implement dimensions of e-health, including electronic medical records, electronic health records, personal health records and e-prescribing – along with expanding initiatives on mobile health in developed countries and emerging technologies – has sparked academic inquiry into the protection of privacy and data and the technology to protect privacy and data. This article examines health information technology policies in the United States and in China and the use of authentication technologies to assess biometrics as privacy's friend or foe in different political frameworks with varying conceptions of privacy. An analysis of privacy in the context of health data protection, challenging relations of trust between patients and providers, the increasing perspective of health data integrity as a cyber-security issue, and the growing rate of medical fraud and medical identity theft may yield findings of a convergence of views of privacy and biometrics unexpected of contrasting political cultures.

*Data security and confidentiality guidelines for HIV, viral hepatitis, sexually transmitted disease, and tuberculosis programs: standards to facilitate sharing and use of surveillance data for public health action. Atlanta, GA: Centers for Disease Control and Prevention; 2011 (http://www.cdc.gov/nchhstp/publications/index.htm).

> A goal of CDC's National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention (NCHHSTP) is to strengthen collaborative work across disease areas and integrate services that are provided by state and local programmes for prevention of HIV/AIDS, viral hepatitis, other sexually transmitted infections and TB. A major barrier to achieving this goal is the lack of standardized data security and confidentiality procedures, which has often been cited as an obstacle to programmes seeking to maximize use of data for public health action and provide integrated and comprehensive services.

Department of Health and Human Services. Standards for privacy of individually identifiable health information: final rule. Fed Regist. 2002;67:53182–273 (http://www.hhs.gov/ocr/privacy/hipaa/understanding/summary/guidanceallsections.pdf).

> The Department of Health and Human Services 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.

*Policy Engagement Network. Electronic health privacy and security in developing countries and humanitarian operations.

London: London School of Economics and Political Science; 2011 (https://www.privacyinternational.org/reports/medical-privacy-and-security-in-developing-countries-and-emergency-situations).

Expertise on the privacy and security aspects of the eHealth systems being deployed in resource-constrained environments such as developing countries and humanitarian operations is severely lacking; the knowledge base in this space is similarly weak. To be effective, the principles and aspirations for medical privacy enshrined in international agreements, policies and commitments must be supported by a local awareness of privacy responsibilities, a strong national legal and regulatory footing, and the appropriate use of information and communications technology. Among the legal and regulatory requirements for strong privacy and security protections are respect for self-determination, the appropriate and proportionate collection, management, access and disclosure of medical information, and strong mechanisms for monitoring compliance and accountability. Any solutions to medical privacy or health information security in these contexts will need to incorporate both technological means such as directed identifiers, access controls and encryption, as well as appropriate organizational, legal and policy responses. Any decision by funders, designers or implementers to exclude these privacy and security mechanisms from an eHealth system must be made as the result of informed deliberation rather than as a matter of expediency. This report was prepared by the Policy Engagement Network for the International Development Research Centre.

Goldman J, Mulligan D. Privacy and health information systems: a guide to protecting patient confidentiality. Washington, DC: Center for Democracy and Technology; 1996 (http://www3.interscience.wiley.com/journal/119099433/abstract).

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.

Interim guidelines on protecting the confidentiality and security of HIV information. Geneva: Joint United Nations Programme on HIV/AIDS and United States President's Emergency Plan for AIDS Relief; 2007 (http://data.unaids.org/pub/Manual/2007/confidentiality_security_interim_guidelines_15may2007_en.pdf).

A 3-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.

Karp DR, Carlin S, Cook-Deegan R, Ford DE, Geller G, Glass DN, et al. Ethical and practical issues associated with aggregating databases. PLoS Med. 2008;5:e190 (http://medicine.plosjournals.org/archive/1549-1676/5/9/pdf/10.1371_journal.pmed.0050190-S.pdf).

We convened a panel of bioethicists, scientists and legal experts to analyse 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. Observations about ImmPort should be relevant to other efforts directed at aggregating databases.

Kruse RL, Ewigman BG, Tremblay GC. The zipper: a method for using personal identifiers to link data while preserving confidentiality. Child Abuse Negl. 2001;25:1241–8 (http://www.sciencedirect.com/science/journal/01452134).

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

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

Results: the research team never learned the child abuse and neglect report status of individuals, the state agency never saw the research data, and the desired analyses were completed using the merged dataset.

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

*Malin B, Karp D, Scheuermann RH. Technical and policy approaches to balancing patient privacy and data sharing in clinical and translational research. J Investig Med. 2010;58:11–18 (http://www.ncbi.nlm.nih.gov/pubmed/20051768).

Introduction: clinical researchers need to share data to support scientific validation and information reuse and to comply with a host of regulations and directives from funders. Various organizations are constructing informatics resources in the form of centralized databases to ensure reuse of data derived from sponsored research. The widespread use of such open databases is contingent on the protection of patient privacy.

Methods: we review privacy-related problems associated with data sharing for clinical research from technical and policy perspectives. We investigate

existing policies for secondary data sharing and privacy requirements in the context of data derived from research and clinical settings. In particular, we focus on policies specified by the United States National Institutes of Health and the Health Insurance Portability and Accountability Act and touch on how these policies are related to current and future use of data stored in public database archives. We address aspects of data privacy and identifiability from a technical, although approachable, perspective and summarize how biomedical databanks can be exploited and seemingly anonymous records can be re-identified using various resources without hacking into secure computer systems.

Results: we highlight which clinical and translational data features, specified in emerging research models, are potentially vulnerable or exploitable. In the process, we recount a recent privacy-related concern associated with the publication of aggregate statistics from pooled genome-wide association studies that have had a significant impact on the data-sharing policies of National Institutes of Health-sponsored databanks.

Conclusion: based on our analysis and observations we provide a list of recommendations that cover various technical, legal and policy mechanisms that open clinical databases can adopt to strengthen data privacy protection as they move towards wider deployment and adoption.

Mizani MA, Baykal N. A software platform to analyse the ethical issues of electronic patient privacy policy: the S3P example. J Med Ethics 2007;33:695–8 (http://jme.bmj.com/cgi/content/abstract/33/12/695).

Paper-based privacy policies fail to resolve the new changes posed by electronic health care. Protecting patient privacy through electronic systems has become a serious concern and is the subject of several recent studies. The shift towards 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 health-care setting. It is designed to simulate different access levels and rights of various professionals involved in health care 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.

Szolovits P, Kohane I. Against simple universal health-care identifiers. J Am Med Inf Assoc. 1994;1:316–19 (http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=116211&blobtype=pdf).

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, 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.

Report on the review of patient-identifiable information. London: United Kingdom Department of Health; 1997 (http://webarchive.nationalarchives.gov.uk/20130107105354/http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_4068403).

> 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 Department of Health/British Medical Association working group looking at National Health Service (NHS) information management and technology security and confidentiality, the chief medical officer established the Caldicott Committee to review the transfer of all patient-identifiable information from NHS organizations 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.

Upshur RE, Morin B, Goel V. The privacy paradox: laying Orwell's ghost to rest. Can Med Assoc J. 2001;165:307–9 (http://www.cmaj.ca/cgi/reprint/165/3/307.pdf).

> 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.

Further reading: implementation and case studies

Frank L. Epidemiology: when an entire country is a cohort. Science 2000;287:2398–9 (http://www.scienceonline.org/cgi/content/summary/287/5462/2398).

> 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 an excellent 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.

Lichtner V, Wilson S, Galliers JR. The challenging nature of patient identifiers: an ethnographic study of patient identification at a London walk-in centre. Health Inform J. 2008;14:131–50 (http://jhi.sagepub.com/cgi/content/abstract/14/2/141).

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, and no sharing or integration of a patient's data, without the retrieval and matching of existing records. Yet often there are 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.

*Nilekani N. Building a foundation for better health: the role of the Aadhaar number. Nat Med J India 2011;24:133–5 (http://nmji.in/archives/Volume-24/Issue-3/Editorial-II.pdf/).

Perhaps one of the most important signs that a country has effectively responded to its governance issues is in the management of its health-care challenges. Health care presents a particularly complex situation for governments, one that is not easy for developing nations to tackle. Health-care goals cannot be reached through funding and budgets alone. Rather, they require a coordinated effort from both the government and a network of service providers; they require programmes and policies that connect with individuals throughout their lives, from birth to death. These requirements are not easy to fulfil in developing nations because of the informal nature of their markets The Aadhaar number offers, for the first time, a tool for governments and policy-makers to knit together disparate solutions to our health-care challenges. While the number is not a single-dose panacea, it is a powerful mechanism for governments and health-care providers to deliver health services more effectively and thoroughly to the Indian resident.

Pedersen CB, Gøtzsche H, Moller JO, Mortensen PB. The Danish civil registration system: a cohort of eight million persons. Dan Med Bull. 2006;53:441–9 (http://www.danmedbul.dk/Dmb_2006/0406/0406-artikler/DMB3816.pdf).

Background: the Danish Civil Registration System (CRS) was established in 1968, when all people alive and living in Denmark were registered. Among many other variables, it includes individual information on personal identification number, sex, 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 people in the CRS, we considered all people registered on 4 November 2005 – that is, all people who were alive and resident in Denmark for at least 1 day from 2 April 1968 to 4 November 2005, or in Greenland from 1 May 1972 to 4 November 2005.

Results: a total of 8 176 097 people were registered. On 4 November 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 people born in Denmark in 1960 or later, the CRS contains complete information on maternal identity. Among people born in Denmark in 1970 or later, the CRS contains complete information on paternal identity. Among women born in Denmark in April 1935 or later, the CRS contains complete information on all their children. Among men born in Denmark in April 1945 or later, the CRS contains complete information on all their children. The CRS contains complete information on immigrations and emigrations from 1971 onwards, permanent residence in a Danish municipality from 1971 onwards, permanent residence in a municipality in Greenland from May 1972 onwards, and full address in Denmark from 1977 onwards.

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.

Rotich JK, Hannan TJ, Smith FE, Bii J, Odero WW, Vu N, et al. Installing and implementing a computer-based patient record system in sub-Saharan Africa: the Mosoriot medical record system. J AMIA. 2003;10:295–303 (http://jamia.bmj.com/content/10/4/295.extract).

We implemented an electronic medical record system in a rural Kenyan health centre. Visit data are recorded on a paper encounter form, eliminating duplicate documentation in multiple clinic logbooks. Data are entered into a Microsoft Access database supported by redundant power systems. The system was initiated in February 2001, and 10 000 visit records were entered for 6190 patients in 6 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.

Tang PC, Ash JS, Bates DW, Overhage JM, Sands DZ. Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption. J AMIA.

2006;13:121–6 (http://www.jamia.org/cgi/reprint/13/2/121).

> Recently there has been a remarkable upsurge in activity surrounding the adoption of personal health record systems for patients and consumers. The biomedical literature does not yet adequately describe the potential capabilities and utility of personal health record systems. In addition, the lack of a proven business case for widespread deployment hinders personal health record 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 personal health record-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 personal health records 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 personal health record systems and provides definitions, system characteristics, technical architectures, benefits, barriers to adoption, and strategies for increasing adoption.

Health identification card implementation guide. Reston, VA: Workgroup for Electronic Data Exchange; 2007 (https://www.wedi.org/knowledge-center/resource-view/resources/2013/02/01/wedi-health-identification-card-implementation-guide-version-1.1-with-errata).

> 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 programmes and others.

*Zelazny F. The evolution of India's UID program: lessons learned and implications for other developing countries. Washington, DC: Center for Global Development; 2012 (policy paper 008; http://www.cgdev.org/content/publications/detail/1426371).

> India has embarked on an ambitious new programme to provide its citizens and residents a unique, official identity. The universal identity programme aims to improve the delivery of government services, reduce fraud and corruption, facilitate robust voting processes, and improve security. It is by far the largest application of biometric identification technology to date and will have far-reaching implications for other developing countries that are looking to adopt national identity programmes to further social and economic development. This paper discusses the evolution of the universal identity programme, the innovative organization and path-breaking technology behind it, how it is being rolled out, and how robust identity is beginning to be used. The paper also draws lessons for other countries. Unlike many legacy national identity programmes, the universal

identity programme is designed from the ground up to support authentication. Its use of multimodal biometrics increases inclusion into the main enrolment database and has a huge impact in improving accuracy. It relies on mobile technology but has also become a driving force behind the development of that technology. Its standards-based approach opens the way for vendor competition and cost reduction. At the same time, its exclusive focus on authentication still leaves the problem of how to validate certain aspects of identity, such as citizenship status. The paper discusses this in the context of the turf war between the universal identity and the national population registry.

Printed on FSC-certified paper

UNAIDS
Joint United Nations
Programme on HIV/AIDS

UNHCR
UNICEF
WFP
UNDP
UNFPA
UNODC
UN WOMEN
ILO
UNESCO
WHO
WORLD BANK

20 Avenue Appia
1211 Geneva 27
Switzerland

+41 22 791 3666
distribution@unaids.org

unaids.org