Implementing National Unique Identifiers for Improving Health or Social Services in Countries: Considerations and Guidance
<table>
<thead>
<tr>
<th>Date</th>
<th>Name</th>
<th>Revision Overview</th>
</tr>
</thead>
<tbody>
<tr>
<td>July 30, 2011</td>
<td>Tom Hutton</td>
<td>Produced initial draft</td>
</tr>
<tr>
<td>October 17, 2011</td>
<td>Tom Hutton</td>
<td>Added additional verbiage to Populating the National Registry; inserted Appendix A – Unique Identifier Systems – Key Concepts</td>
</tr>
<tr>
<td>November 22, 2011</td>
<td>Eddy Beck, Tom Hutton, Linda Mattock, Jan Macgregor, Xen Santas, Steven Uggowitz, Meade Morgan</td>
<td>Revised introduction, substantial edits and clarifications, some revision to organization and structure, added Appendix E</td>
</tr>
<tr>
<td>June 1, 2012</td>
<td>Tom Hutton</td>
<td>Implementation Models and Resource Estimation Appendix added.</td>
</tr>
</tbody>
</table>
1 Introduction

Many middle- and lower-income countries are in the process of scaling up HIV and other medical and social services, resulting in the collection of increasing amounts of individual-associated data. Given these increases in the amount of data collected, stored and tied directly to unique individuals a need exists to ensure that the confidentiality and security of such personal data are protected\(^1\). Many countries are now developing unique identifiers for their citizens within the context of establishing national health standards and implementing systems for storing and exchanging information about health services. The need to uniquely identify patients receiving various HIV or other services is a logical byproduct of using modern information management systems in the context of longitudinal patient care. At national level it is becoming increasingly critical for service provision and programmatic planning both to monitor outcomes of those receiving services, to monitor population impact and to ensure complete coverage and adequate resources in terms of personnel and supplies. This includes the ability to uniquely identify persons using these services over time as well as across service delivery sites and the ability to effectively store and exchange the information at and between local service delivery sites, regional registries or data-warehouses and national ministry of health offices or other national authorities.

Depending on the available technology, policy environment and complementary efforts both within the health sector and other government entities, there are several methods available to uniquely identify individuals in a nationally scalable and deployable way. Each has its associated objectives, advantages and limitations. All methods need to ensure that the confidentiality and security of such personal information is protected but the most appropriate methods for each country can only be determined by weighing strengths and challenges against the specific needs of the country and its population.

To assist this process, UNAIDS with PEPFAR (President’s Emergency Plan for AIDS Relief) organized a workshop in February 2009 bringing together a number of different countries that were in the process of developing unique identifiers for health or other services\(^2\). Discussion during the workshop focused primarily on the landscape of challenges and benefits of national level individual identifiers standards, as well as specific designs for a health identifier and their implementations. This workshop also concluded that the development of a sound health information system is important to providing reliable information for planning and managing HIV or other services\(^1\). Participants also agreed that the development and implementation of such unique identifiers should be complemented by the development and implementation of guidelines on protecting the confidentiality and security of personal information\(^2\).

This document describes guidelines and considerations for national implementation of individual identifiers. Because such identifiers typically exist within the context of national information systems, it also addresses national standards for data storage, exchange and touches on associated technology, policy and sociological considerations important in any deployment. The options are presented at a conceptual level to inform readers on scope of such systems. To implement them, each country must consider implications within their national context and determine appropriate implementations for their specific needs. This document is meant to provide only a starting point to aid in highlighting the many and diverse functional areas that need to be considered to deliver and sustain such projects successfully.

\(^1\) Guidelines on Protecting the Confidentiality and Security of HIV Information: Proceedings from a Workshop

\(^2\) Developing and Using Individual Identifiers for the Provision of Health Services including HIV
2 Patient identification in healthcare settings

2.1 Overview

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

1. **Coordination of Patient Care Services** - Interacting with other service domains such as laboratory, x-ray, dietary, physical therapy, etc. and communicate orders, results, request for services, supplies, consultation etc.
2. **Record Keeping/Information Management** - Collecting and organizing information such as orders, results, procedures, notes, etc. into a manual medical record chart or in an automated electronic medical record for current and future use.
3. **Administrative Functions** - Handling administrative functions including billing and reimbursement.
4. **Storage and Retrieval of Historical Information** - Retrieving and reviewing past medical history including problems, diagnosis, procedures, medication, allergy, etc.
5. **Aggregation of information from multiple patient information** - Collecting, aggregating and performing analysis on groups of patients for treatment efficacy, research, statistical reporting and planning.

In addition, patients may be mobile, visit multiple providers and be treated by multiple and different types of organizations even across broad geographical areas. Lack of a unique patient identifier presents significant problems in maintaining continuum of care within facilities and across multiple providers where access to information from multiple care settings and the retrieval and assembly of relevant patient care information from past episodes of care affects the delivery of care. The unique patient identifier is needed when:

1. aggregating and providing a record of a patient's information;
2. accessing and integrating information from different providers and provider computer systems;
3. supporting population-based research and development.

Therefore, supporting continuum of care is one of the compelling reasons to uniquely identify patients across multiple providers and locations.

Country-wide systems designed to precisely identify individuals and events of interest, such 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. This may artificially inflate the amount of disease or exaggerate the severity of its impact.

**Use of universal patient identification allows:**
- The ability to assemble a longitudinal patient record for more informed and accurate treatment.
- The ability to accurately de-duplicate aggregate information.
- More accurate data quality assessments and higher data quality.
- More accurate aggregate (indicator) reporting for program management.
- Enhanced avoidance of fraud with drug stocks.
- Enhanced research accuracy.

---

1 From: [http://www.ncvhs.hhs.gov/app2.htm](http://www.ncvhs.hhs.gov/app2.htm)
2.2 The Value of Having Accurate Universal Patient Identification

Current methods for patient identification almost always involve the use of a medical record number\(^2\), issued and maintained by the provider organization. Often, this number is based on an institutional Master Patient Index (MPI) 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 effort made to ensure identifier uniqueness. As a result, patients often become linked to several medical record numbers, each issued by the clinic or hospital that provided them care. These numbers provide unique identification only within the issuing organization or organization location. 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 those instances where the same medical record number was issued in more than one clinic, facility, or organization, distinct individuals may be erroneously considered to be the same person because they share the same medical record number.

2.3 Substitutes to Coordinated Approaches

In the absence of a dedicated and coordinated approach at national level, a variety of approaches have been devised as substitutes or surrogates for nation-wide systems.

For example, one approach to uniquely identify individuals across clinic, clinic sections or locations is to extend medical record numbers by assembling a set of identifying characteristics and using 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 falling within an acceptable probability of being the same individual are treated as the same individual and those outside the range are treated as distinct individuals. The set of characteristics and their associated weights should be specific to the cultural context of the country. For example, one set of identifying characteristics might be name, gender, age, race and address. Another set of characteristics might be gender, age and tribal affiliation.

While used in multiple countries, this method is limited by the accuracy and completeness of the captured identifying characteristics. In addition, because this method relies on identifying characteristics, it requires measures to ensure the confidentiality of the individuals.

Most importantly, any method system where individuals cannot be identified with 100% certainly should not be used for direct patient management. Instead data from systems is generally best used for only for aggregate reporting and statistical analysis.

2.4 The Various Levels of Patient Identifier Usage\(^3\)

Where and how a unique identifier will be used often plays an important role in how the identifier is designed. The design in turn may either limit or facilitate the use of the identifier beyond the original setting.

---

\(^2\) This may in fact be an alphanumeric sequence. The term ‘number’ is used here for simplicity.

\(^3\) Loosely based on: [http://www.ncvhs.hhs.gov/app2.htm](http://www.ncvhs.hhs.gov/app2.htm)
The identifier may be used at different levels within a single organization or across multiple organizations. For reference it is important to be able to identify specifically at which levels patient identifiers are being used. This is called the “scope” of the identifier. In general the scope can be divided into the six (6) categories or levels listed below (All five categories of uses described earlier are applicable to all six levels):

1. **Facility section use** – At the lowest level, the identifier is only unique within a particular section of a health facility, typically used in a paper record such as a logbook.
2. **Facility level use** – At the second level, the identifier is typically used for a patient folder, which is utilized across facility sections.
3. **Organization-wide use** - At the third level, the identifier's scope of use is limited to functions within the provider organization. The current use of patient identifiers by most healthcare organizations is at this level.
4. **Enterprise-wide use** - At the fourth level, the identifier's scope of use includes an entire enterprise. The enterprise may include multiple provider organizations providing same or different types of services. The identifier in this case is used to identify an individual and provide enterprise-wide access to his or her medical record/information. The patient has access to care across the enterprise at this level.
5. **Nation-wide use** - At this level, the identifier's scope is expanded for nation-wide use among healthcare organizations. The full potential of a patient identifier can be realized at this level: 1) access to and use of patient care information from different providers for the purpose of delivery of care, 2) electronic integration of information from different providers, 3) lifelong view of a patient's information and 4) aggregation of population-based information for research and development. The national healthcare reform initiatives, such as managed care and integrated delivery networks, have expanded the patient identifier's scope to a nation-wide use.
6. **Global-use** - At this level, the identifier's scope is further expanded to world-wide use. All benefits and uses discussed earlier are transformed to global level.

### 3 Health Identifiers Used Nation-Wide

#### 3.1 Why a (nation-wide) National Health Identifier?

To truly take advantage of information systems and information communication and technology (ICT) in the areas of health statistics, monitoring and evaluation, surveillance and longitudinal patient care, a centrally planned and coordinated approach to unique patient identifiers is really the only option. This approach: to implement a National Health Identifier (NHID) is the focus of this document and associated efforts.

With an NHID, a country builds and runs a system that can assign and manage a single unique number to each individual. This number 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, or sex.

#### 3.2 Defining a National Unique Identifier

If a country intends to go through the process of implementing a nation-wide identifier system, it is important that the appropriate department within the government address the important policy, design and management questions that are always part of such projects.
It might be that the effort needs to involve more than one department however. Questions that needed to be asked in order to find the proper department or set of departments for the effort include the following. Who in the government is responsible for the definition, methods, and issuance of identifiers to people? Who has the legal authority and the moral and popular support to lead such an effort? Are there sufficient government organizations to oversee this effort? In addition, are there sufficient legal protections to protect misuse of patient information and provide accountability for the use of the information? Can the infrastructure of the system henceforth be deployed and maintained in a robust and secure manner that meets the requirements.

In the lead up to any NHID project, the following fact-gathering activities should be considered:

1. Make an inventory of methods for identifying individuals and their health records already in use in the country, including those implemented by non-government organizations (NGOs) and other entities that are not within the country’s public healthcare system but for whom the NHID would still be applicable.
2. List the needs that the NHID is expected to meet. These might include (1) short-term for individual identifier, family linkages, disease and disaster control, medical statistics and (2) long-term for keeping healthcare records.
3. Calculate the cost and resources needed to provide accurate individual identification and the trade-offs among different options.
4. Develop a plan to implement the new NHID numbers, including how it will be announced to the medical community, government officials, NGOs, and software developers to allow these groups sufficient time to plan for the use of the new number.
5. Evaluate the existence and implementation of a country’s guidelines on the protection of the confidentiality and security of personal information.
6. Evaluate the ability of all responsible institutions touched by the NHID project to plan, deploy and oversee ongoing operations of the their piece(s) of the work.

In addition to considering the management questions and completing the preparation activities, each country will need to determine the scope of the NHID, how deployment will be scaled-up and the contexts in which the identifier will be used.

4 Card-based National Health Identifiers

4.1 Determining the Scope of the National Identification Cards

For countries that already use unique identifiers in other facets of government (civil registration, taxation, etc), at first glance it may be easy to conclude that the easiest or best option would be to scale up the use of such existing systems for use in the health domain. Unfortunately making such a decision is generally not that simple. There are many diverse considerations that can impact such a decision and they can vary greatly from context to context.

As an example, let’s consider the utility of driver’s licenses and passports and note that there is a difference between an apparent identification and an actual identification. While drivers’ licenses and passports are treated as actual identification, they can be forged, even to the extent that there are many people who make their living making phony identifiers. Not all users of the health care system may be eligible for such cards either.

Generally, the larger the scope and use of the identification card, the larger the motivation to defraud the government or a citizen, and when doing so, threaten the integrity of the data system to the extent that the forged number at some time was assigned by the government to another individual who is also in the system. Therefore, while a national universal identity card could conceivably be utilized for all
identification needs, caution should be exercised in assuming that a single identifier should be universally used to identify someone for everything, particularly for financial transactions and police matters.

In addition, there are specific confidentiality considerations having to do with a patient’s medical data, such as how the data is used and disseminated.

All technological solutions entail risks, which should be identified, evaluated, and discussed fully before adoption. For instance “smart cards” when used as NHID 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, some persons may do so, and as a result the public could actually be less safe by using the card.

In addition, universal identifiers can be misused by authorized inside users or unauthorized outside users. There are many documented examples worldwide of abuse of the personal information stored in databases, such as privacy abuses, overzealous surveillance, and personal information being used to locate, stalk, or even murder persons. Because of the possibility or history of extensive abuse of a universal identifier, there are some countries that constitutionally or legislatively severely curtail or forbid the use of a universal identifier.

Therefore, it is important for each country to carefully consider what the use of the identifier should be and, if reasonable, whether it could be limited 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 an “as needed basis”.

### 4.2 Planning the Issuance of the National Health Identification Cards

Implementing a NHID and operating an associated card distribution system (creation, issuing, accessing, updating, invalidating/deleting data) requires careful strategic planning. The plan itself 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. The following is a non-exhaustive list of some of the more important considerations:

1. Planned starting date for issuing cards;
2. Planned time frame to issue cards to 90% of the target population;
3. Projected intermediate dates for 30%, 50%, and 75% coverage levels;
4. Total cost of ownership (TCO) and lifecycle planning;
5. Incentives and links to other identification efforts in the country incl. private sector, insurance providers, civil registration etc.
6. Earliest issuance in the life cycle of a person – during pregnancy versus at birth versus after birth;
7. Geographic and age distribution patterns of the citizens.
8. Where are identity cards and other documents issued, including birth certificates, voting cards, citizen certificates or passports, drivers and other types of transportation licenses, postal offices, government locations issuing business licenses, community centers, hospital and clinic reception points, induction into the military, and employment application points?
9. Should there be temporary locations established that can issue the card?
10. Will biometric identifiers be used in the medical and possibly other systems to identify persons?
11. Will pictures be taken of the applicant?
12. Will other biometric information of the applicant be gathered?
13. What type of information will be gathered at the point where the identifier is issued?
14. Will biometric identifiers be stored on the card?
15. Where else should the card be issued, such as at hospital outpatient and inpatient reception points, clinics or other sites?
16. Will the cards be issued immediately, or will there be a period of time in which the application is reviewed and approved?
17. When the applications are approved, will all points of where identifiers are issued have the ability to generate the card?
18. How many people will be required to staff the points of application and issuance?
19. Roles, responsibilities, security and privacy considerations of the NHID related workforce
20. What training is required to train the persons interviewing the patient, issuing cards, and support this effort?
21. How robust will the information and communication technology (ICT) system need to be if points of application or issuance are required to be open to the public at all times? Who will provide technical support at the application or issuance points?
22. To protect against NHID cards being issued fraudulently, some sort of verification process should be formulated. The verification process will need to consider the following:
   - any documents not in the national official language will need to be translated by a certified translator;
   - depending on the country and varying local conditions in villages or tribal locations, there may be a minimum of formal documentation in existence which would help verify a person’s identity;
   - alternative identification methods may have to be developed, such as meeting with a village or tribal elder or other leader; village or community healthcare worker; religious leader; or other source that is considered trustworthy, in order to ensure that the application has integrity; and,
   - the costs of all of the above items, financially and in human resources.

Documents are not always of equal value in verifying a person’s identity. One approach is to use a scoring system build a pass or fail test for the identifying documents. Such scoring systems typically contain four or more levels:
   - The first level documents with the highest credibility,
   - The second level documents are still relatively high in credibility, but do not have the same level of credibility as the first level,
   - The third level documents contain less credibility than levels one and two and
   - The fourth level documents that have only limited credibility.

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

Document examples for the first level might include:
   - Records from or issued by health facilities (Facility ID Cards, immunization cards etc.)
   - Birth Certificate or birth card; original or certified copy
   - Citizenship Certificate
   - Current Passport
   - Expired passport that was valid within the preceding 2 years
   - Other identity documents of the same characteristics as a passport include diplomatic documents and some documents issued to refugees such as Consular Report of Birth Abroad, Certificate of Naturalization, and court or marriage or divorce documents that provide proof of a change in name that differs from the primary document presented.

Second level documents might include:
   - Driver license issued within the country;
   - Government Photo Card;
   - License or permit issued under a law of the Commonwealth, a State or Territory Government: for example a boat license;
   - Identification card issued to a government employee;
   - Identification card issued by the Commonwealth, a State or Territory Government as evidence of the person's entitlement to a financial benefit;
   - An identification card issued to a student at a tertiary education institution.
Third level documents, all of which must be signed or consented by the applicant or by a competent and trusted person witnessing the oral understanding and agreement, might include:
- A document held by a cash dealer giving security over the applicant's property;
- A mortgage or other instrument of security held by a financial body;
- A document from your current employer or previous employer within the last 2 years;
- A land title record;
- A marriage certificate (often useful to determine maiden name and names of parents);
- A credit card;
- A foreign driver license;
- A government issued card from another nation, which does not have a picture.

Fourth level documents might include:
- A record from a public utility, such as a telephone, water, gas or electricity bill;
- A record from a financial institution;
- An Electoral Roll compiled and verified by the national voting authorities;
- A record held under a law other than a law relating to land titles;
- A lease or rent agreement;
- A rent receipt from a licensed real estate agent;
- A record of a primary, secondary or tertiary education institution attended by you within the last 10 years;
- A record of professional or trade association of which you are a member.

An example scoring system could require a minimum of 100 points. Each document from the first level could be scored at 100 points, and therefore be authoritative on their own. Documents from the second level could be scored at 75 points each, thus requiring two of them or a mix of other documents. Each document from the third level could score 40 points, and documents from the fourth level could be worth 20 points.

It is important to consider all of these documents, as well as other authoritative documents that could be considered, within the national context and evaluate each type of document and what value it could have in this process.

5 Storing and Managing the Data for Nation-Wide Systems

5.1 Design Considerations and Populating the National Registry

Unique identifier systems can be designed and built in many ways. The chosen path should be decided based on a combination of:
1. the overall system objectives,
2. the policy environment; and
3. the on-the-ground realities of what systems may realistically be deployed and sustainably operated.

5.1.1 Where should data be stored?

A nation-wide system does not necessarily need to have all the data hosted or managed centrally. There are actually quite a few approaches that can be used instead of the most natural assumption that all the data will be stored in one big computer in one location. For example the American Society for Testing and Materials (ASTM – http://www.astm.org) has described several alternative approaches such as:

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 location or patient identification information preceded by a non-unique patient number in order to attempt to make the patient number unique, there are several significant flaws in this method. First, as regions and districts 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 then becomes, do you re-issue the patient number and re-number all existing patient data, or do you have patient numbers with area codes which are not correct? If you mix patient identifiers, such as birth date, sex, or some other identifier with the number, you will be constantly disclosing personal information every time the patient identifier is used. Also, if the birth date were to be incorrectly recorded and a correction needs to be applied, now the NHID has been changed, so all existing electronic and paper medical records would have to be changed as well. For these reasons, a method which mixes data, such as location or personal identifiers, with an “only unique at one location” is not desirable.

2. **Centralized national unique identifier assignment for multiple networked locations.** The scenario proceeds as the first method, assuming that all medium or larger facilities are semi-reliably networked and therefore can share data on a semi-real time basis, within 30 minutes. While this approach is highly desirable and helps avoid duplicating identities, it may not be possible at all locations, and may be unreliable or costly as well. This method can be combined with the fourth method, a hybrid approach of 70% or more of the patient volume being interconnected and 30% done in timely batch is considered a reasonable compromise in many circumstances.

3. **Hybrid centralized national unique identifier assignment for multiple networked and semi-networked locations.** This approach is the one mentioned in the second method plus some locations using the forth method. It is important that all of the sites that are considered medium to high patient volume are interconnected in order for this approach to function reasonably well. It is also desirable for as many as possible lower volume clinics to also be networked. While the data volume between busier clinics in the same region will be higher, the data volume of the remote clinics should be fairly small. A good result may be achievable with daily cell phone data transfer sessions or a memory stick being sent to the closest networked clinic or hospital.

4. **Centralized national unique identifier assignment for multiple disconnected locations.** Identifier uniqueness and ultimately overall system success relies 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 point of central assignment, prompt sending and receipt of a request electronically by phone, memory stick, or on paper. However, paper should be discouraged if at all possible. The process may further depend on centrally produced, serialized identifier cards being pre-distributed, although that should be discouraged as possible as was discussed in method one. Across an entire or the majority of a country, this approach often works poorly and is undesirable.

5. **Locally assigned national unique identifier assignment for multiple disconnected locations.** All sites are assigned a unique provider-site code - 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 patient in durable, portable form and is associated with all subsequent encounters and data deriving from that.

---

4 Note this language refers to real-time IT database systems. In most developing countries encounter/transaction volume and bandwidth considerations will be different because they may be batched (overnight/daily etc.)
5.1.2 Data Linked to the NHID and Interoperability with Other Information Systems

Typically the information populating the national registry will need to be obtained from a number of places, including birth, and death databases; points of application for and issuing of the NHID card; and patient service locations. In order to get information from the various sources, both batch-oriented and interactive software will have to be developed. In many contexts paper based systems may also need to play a large role, especially at the lowest levels of the health system. In addition, there will need to be a well-thought-out process for resolving information that seems at odds with each other.

If family relationships are stored in the database, the information will need to also be related in family units within the database and appropriate sources for family unit information identified. These sources typically include birth and death records, and citizenship and passport applications.

A NHID will be assigned either in the absence of a pre-existing healthcare identifier or it will be deployed in the presence of an existing health facility patient identifier, or multiple identifiers, which is more likely to be the case. It is also likely that any identifiers that exist are only unique at each patient service delivery facility or facility section.

5.1.3 Deploying and Integrating the NHID at Points of Service Delivery

In the case that a healthcare identifier already exists in a given place of service delivery, the existing systems - paper or electronic - must be altered to accommodate the new NHID. This modification may be time consuming and expensive, since it must include mapping all of the historical data to the new universal healthcare identifier. In addition, issuing of NHIDs, and capturing and mapping the new unique healthcare identifier may introduce new ways of working. For example, a pre-existing patient identifier that is internally generated 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 that patient identifier to each distinct healthcare entity. Existing technology will likely need to change 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 will vary across different healthcare settings because existing processes for generating and maintaining healthcare identifiers are highly variable.

Assigning and managing national identifiers may be conducted either by a central assigning authority or may be distributed and coordinated among many loosely connected or disconnected sites. While each approach has advantages and disadvantages, real-world constraints may ultimately dictate the development of a particular architecture. For example, unreliable or nonexistent network connectivity in resource constrained settings may dictate that the assignment of unique identifiers be distributed among many loosely connected or disconnected sites. 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 healthcare identifier can improve patient identity management, it is not a panacea for a variety of reasons. Sophisticated matching methods are a necessary component of any robust unique identifier system. Selection of appropriate matching methods is addressed in other sources.

From a technical perspective, comprehensive patient matching methods will still be needed for the following non-exhaustive list of functionalities:

- **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 healthcare identifier system ideally introduces no duplicate identifiers. That is, a NHID is only assigned to a single patient; two or more patients should never share the same unique identifier. However, duplicates may arise for a variety of reasons and therefore a system
for managing unique healthcare identifiers must implement a process for reconciling such duplicates. The process for doing so may follow this general pattern:

- Duplicate identifier is detected through a potential variety of automated or manual processes.
- Duplicate identifier is deactivated; no further use of the identifier is allowed.
- New unique identifiers are assigned to all individuals who shared the single identifier.

- **Incorporating historical data.** To fully leverage the benefit of a NHID, clinical data collected prior to implementing a NHID system must be retroactively linked to the unique identifier. To do so requires mapping demographics and old identifying information to the new unique identifier. Depending on the desired result, this will may be accomplished minimally by linkage within a national registry, or by re-labeling all paper records. The large amount of work required to implement a comprehensive renumbering of all patient records will require extensive resources, and therefore may not be practical.

- **Identity theft and the sharing of identifiers also pose challenges.** It is difficult to detect medical identity theft or identifier sharing unless a provider notices a discrepancy in the clinical data; for instance some records indicate that a person has diabetes, while the others do not. The use of someone else’s information need not be criminal or malicious; it may be as simple as parents’ providing their 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 healthcare identifier, the identity of an individual claiming the unique identifier can be confirmed using supplemental identifying information, for instance area where the unique identifier was initially assigned. This information is then matched against other records to ensure an appropriate match.

- **Verifying absence or presence of an identifier.** 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. Prior to 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.

There will also need to be a suitable written registration system at the sites where health care is provided. This system should application that capture the name, any aliases, any 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 input. Any program should be able to also perform basic reporting and label printing. The program will likely need to be deployed at least at the following locations:

1. Hospital outpatient and inpatient receptions;
2. Preventing Mother-to-Child Transmission of HIV clinics (PMTCT), Voluntary Counseling and Testing (VCT) sites, Antiretroviral, (ARV) clinics or Tuberculosis (TB) clinics or wards if patients bypass centralized reception points and go directly to those locations or other clinics;

Where infrastructure and resources at the service delivery site permit there should need to be a computer programs developed that identify and allow editing of information in order to resolve data discrepancies, conflicts, or fraud that may arise. These programs will need to match any electronic records at the site based information like first name, last name, cell phone number, age, birth date, sex, or country of birth; search for multiple records with similar identification characteristics; and editing any and all information. This should include a status field which indicates whether the record is active, inactive, in review, or denied. If the status of the record is “unusable” or “denied”, it would be helpful to also store the reason for this, such as multiple identities being issued to a single person.

In order to facilitate this process, there will need to be a national level group of persons, who are well trained in record matching and identification and data resolution. In addition, there will also need to be a
record status of “pending resolution” that is 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, which allows sorting by date of application or pending status.

The number of persons that will be required to perform this function depends on the rate at which applications and registrations occur per month, and the period of time that resolution should occur. For example, if there are 100,000 applications being received per month, and there are possible discrepancies identified in 5% of the applications, there will be 5000 applications per month that will need to be evaluated and resolved. If the average resolution takes 15 minutes, and assuming an 8 hour work day, each person can resolve about 32 records per day or 160 per week. If the acceptable time to resolve a problem is targeted to be a week, this means that in an average week there will be 125 applications per week that need to be resolved, which is within the capabilities of ten people resolving problems, as per this example.

Please note that these numbers are simply an example, and should not be assumed to be applicable within the setting of any particular country. To determine the amount of time that it really takes to provide resolution to registry conflicts, the actual data being used must be known, and a resolution strategy and a set of scripts for resolving the data questions must be generated and tested. By doing this, the amount of time that is actually needed will be determined, along with other logistical requirements, such as access to a telephone or other communication means, that are required to resolve problems. Also, it should be noted that in any situation, there will need to be a tiered response system in which quick resolution can occur on many records by persons that are authorized to resolve simpler conflicts while a second tier “supervisory” person that investigates more difficult or complex identification issues.

5.2 Identifying and Labeling Existing Patient Records

There are two methods of implementing the NHID at a point of service, the first being just to identify the patient, and the second is to be fully intermeshed at the clinic.

The first method is implemented by using the health identification card to identify the patient, through the use of the name, age, picture, biometrics, or other methods. Once the patient is identified, the already existing local clinic identifier is used to identify the patients. To ensure minimal simple human errors occurring, the patient folders might have the NHID printed a label and attached during the first encounter of a new NHID enrollee. This label is encouraged to have at least a bar code printed with the number, in order to make it easy to identify a patient, and then scan the cover of the patient folder in order to electronically verify that the patient folder and patient who is presenting are known as the same in the system.

The second method is to utilize the NHID in conjunction with all of the existing patient paper records. This method is more costly, but will help assure that manual transcription errors do not occur. To facilitate this, a patient ID area on the forms must be available to write the information, or patient ID labels should be printed and attached to existing paperwork. Therefore existing forms, folders, and other items will need to be evaluated in order to determine the exact label size(s) that are appropriate to utilize.

Once the desired label sizes and formats are known, existing programs that perform registration or other patient functions will need to have a function that provides for printing the required labels on demand in order to further the process of placing the new NHID on all paper records (either folders or folders and patient records). Printed labels can be important not only for records kept within a site but could also be used to assure accurate identification when sending biological specimens for laboratory testing.

Current real-world processes for registering and identifying patients are highly variable across different healthcare settings. Consequently, the processes for assigning unique healthcare identifiers will be variable as well. Despite this variation, all processes will need to include the following common actions for assigning unique identifiers:
• **Initiate request for new identifier.** A new identifier request is initiated by an authorized local person, including the patient, a clinician or an administrator.

• **Provide identifying traits.** To verify the identity of the person for whom the identifier will be created, sufficiently identifying patient characteristics are provided, including name, birth date, gender and others.

• **Identity verification.** Using the supplied identifying information for the particular patient, the local person responsible for issuing them verifies that no known unique identifier has been previously assigned to this patient.

• **Assign new identifier.** The trusted authority generates a new unique identifier in response to a 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.

See appendix C and D for further discussions of the logical and electronic issuance of the NHID.

### 5.3 Functions of a National Health Identifier

The five basic functions that a NHID must support are:

1. Positive identification of the individual in order to receive care or for administrative functions.
2. Identification of information to aid a more informed delivery of care and building a coordinated multi-disciplinary 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 future.
3. Aggregation of information across institutional boundaries for population-based research and planning.
4. Support the protection of privacy and confidentiality through accurate identification through explicit identification of patient information and de-identification once the information leaves the primary site through masking, encryption or anonymization.
5. Reduce healthcare operational cost and enhance the health status of the nation by supporting both automated and manual patient record management, access to care, and information sharing.

The identification number itself should not contain any data such as site identifiers, regional identifiers or any personal identifiers (such as initials or dates of birth.) The reason for not putting these types of data in the NHIDs that any codes that are in use in the country are likely to be modified over time. Once the change happens, all cards issued to date will likely need interpretation or will be invalid, neither of which is a good result. The vast majority of all coding systems will change over a period of a hundred or more years. For example, regional divisions and numbering systems will be changed as populations shift and grow.

The reason that the identification number should not contain any personal identifiers is that any use of the card or the identification number immediately gives away some amount of personal information, the result being the weakening of patient confidentiality. And although it is commonly thought that this data will never change, human errors are made in applications or entering the applicants information into data systems, that will require future corrections; all of these will then invalidate the identification number issued to the person.

### 5.4 Required Components of National Health Identifier and National Patient Registry

A NHID must include linkage components that will provide it with needed functional capabilities. Each identifier must be supported by demographic information for the individual it identifies. Such information must be updated regularly in order to be accurate and usable. The identification process includes
searching Master Patient Indexes (MPIs), matching identifiers and verifying information. Depending on the identifier’s scope and level of use, the search processes can range from a single provider organization to the entire national healthcare system with the possibility of future worldwide expansion. Therefore, the NHID requires a robust technical and administrative infrastructure.

The following seven components are integral parts of the National Patient Identification system. They must work together in order for it to perform its functions and fulfill its objectives:

1. An identifier scheme that can be numeric or alphanumeric in nature;
2. Identification information;
3. Cross-references to local site-specific patient identifiers for existing patient identification numbers;
4. Mechanisms to hide or encrypt the identifiers;
5. Software to mass-register patients, and accompanying personnel to do so.
6. Software to search, identify, match, encrypt or other ways to manipulate the underlying information.
7. Administrative infrastructure including the central governing authority.

6 Confidentiality and Security Considerations

Increasing amounts of individual level data are being collected in middle- and lower-income countries which are scaling up HIV prevention, treatment, care, and support services. While substantial background material and experiences from upper-income countries is available, relatively little advice exists for lower- or middle-income countries to develop and implement their own guidelines in this area. As was recently demonstrated, few countries have developed comprehensive guidelines on protecting the confidentiality and security of HIV information. A workshop was therefore organized by UNAIDS and PEPFAR in May 2006 to develop a set of general guidelines, which could be used by countries as a starting point to develop their own guidelines, and published as the Interim Guidelines on Protecting the Confidentiality and Security of HIV Information. The National Health Identifier Enhances Confidentiality and Security.

Assuring the privacy and confidentiality of patient care information is a difficult challenge and it should not be ignored. As described above, the development of unique identifiers needs to be placed within the broader context of countries developing and implementing guidelines to protect the confidentiality and security of personal information. A NHID is an integral part of the patient care information and it requires the same confidentiality and security protection as the patient care information itself. In fact, the NHID can help meet confidentiality requirements by standardizing and strengthening access control, and eliminating the repeated use of other significant amounts of personal identification information every time access to the patient files occurs.

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 which 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 personal information is provided in the Guidelines on Protecting the Confidentiality and Security of HIV Information: Proceedings from a Workshop.

---

http://www.jiasociety.org/content/pdf/1758-2652-14-6.pdf
The systems supporting the use of the NHID require design architectures that will keep the identification of patient care information and its access as two distinct and separate functions within healthcare. The identifier's role is limited merely to identify the patient record by accessing only the identification segment of the patient record and not its content. Access control deals with the authentication of the user for example the validation of user identifier and password, verification of access privileges, audit trails or physical security. Access control must be supplemented by organizational policies and procedures, as well as 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. The 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.

### 6.1 Role-Based Access Security Ensures Confidentiality, Not the Identifier

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 the authorized use by accurately identifying the patient and his/her information.

A NHID alone will not address the patient identification need. It cannot protect the privacy and confidentiality of patient care information or assure 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 American Society for Testing and Materials (ASTM) International characteristics listed on the next page, such as assignable, accessible, identifiable and others,. deal with compliance by the Issuing Authority, healthcare information is created, maintained, accessed, and used at healthcare organizations. Positive identification of individuals and proper controls on access to their information are required at the healthcare sites.

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 Trusted Authority and 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 is 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 the NHID. The following are examples of measures that can be implemented by organizations that generate, access, and use individual level information:

- Access Protection;
- User Authentication;
- Audit Trails;
- Training & Education;
- Physical Security;
- Organizational Policies and Procedures;
- Promoting Organizational Culture that is conducive to the protection of privacy;
- The use of data classification in determining system security measures;
- Built-in computer hardware & 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.

7 Nation-wide Coordination and Infrastructure

7.1 The National Health Identifier Requires a National Issuing Authority and National Coordination

The issue and maintenance of the NHID, the identification information and their use need to be handled either under a centralized or decentralized administration. This will require a national-level system, often referred to as a patient registry, which would allow authorized persons to add and edit data in order to build a master patient index database. The national system should be distributed into the appropriate regional or district locations to enable access to the patient indexes to the local providers of care and services. Updates from local instances of the data would flow to the national system, and then be synchronized with other local instances of the database as appropriate. The technical issues around distributing and synchronizing these data can be quite complex, and are beyond the scope of this document.

7.2 Ministerial Infrastructure

The ability of a country’s government to develop, deploy and operate a high-quality unique identifier system will depend on a minimum of:

- 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 a minimum of a five year commitment to this effort.

Since the information in a national any national data repository or data warehouse could be easily used for purposes that were not originally intended, either on its own or combined with other information, a well-executed umbrella of governance and infrastructure should be in place from the start of the NHID registry.

A project of this magnitude will require a dedicated team of professionals, which consists of a minimum of:

1. A senior project manager - A person who keeps track of the project either from inception to deployment. Project managers have varying levels of responsibilities and authority
2. Systems Analyst - The person responsible for the development of an information system. Systems analysts design and modify systems by turning user requirements into a set of functional specifications, which are the blueprint of the system.
3. Database Designer - An expert in relational database modeling is essential to the success of any such endeavor.
4. Policy Expert
5. Technology Architect - A senior technology architect with knowledge of information and communication technology (ICT), as well as a broad understanding of data architecture and software development.
6. IT technicians available at each facility where improvements/enhancements will be needed;
7. Technology Committee - A technical working committee, consisting of representatives of the major stakeholders and technical experts, to steer the progress, evaluate work plans, and make decisions on the project steps and other related matters as needed.

Project management frameworks such as PMI or PRINCE2 are of much use to organize the work. Also frameworks for IT service management such as ITIL give useful guidance on all aspects of planning and operating large systems.

8 Policy and Governance

8.1 Governance Mandate

The mandate for governance should be achieved through the passage of a USA Privacy Act or similar acts that:

1. Establishes that the definition of personal information includes written, spoken, electronic, and all other forms of information about an identifiable individual.
2. Recognizes that collection and use of personal information by the government has the opportunity for misuse in ways that are specific to government functions, such as using medical data for criminal investigations.
3. Establishes that the national judiciary is granted the authority to review all claims of denial of access to personal information held by the government; and improper collection, use, and disclosure of personal information.
4. Establishes that the rights of access, correction, and notation with respect to personal information held by a government institution applies to all persons whether citizens of the country or not.
5. Establishes that government institutions can only collect personal information that is reasonable and necessary 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 individual data.
6. In the case of a disclosure of personal information, which 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.
7. Establishes that all data use must be fully disclosed to the individual. The framework established for this should be based on maximizing the transparency of government and ensuring a maximum amount of accountability of government to its citizens.
8. Establishes that no data can be used for purposes that are outside of the intent disclosed to the individual. No data matching, linkages, or aggregation of the data by external systems is allowed except as disclosed in writing to the persons to whom the data belongs.
9. Establishes that all data system usage will be governed by the principle of least privilege. i.e. that people who have access to the system are assigned the least level of access rights necessary to do their jobs.
10. Establishes that the government institution remains accountable for personal information where decisions are made to outsource departmental work, and that the information is considered under the control of the government institution at all times.
11. Establishes and empowers the Office of Information Security or similar government office, which is responsible for:
a. Establishing standards and requirements for protection of personal and non-personal data;
b. Establishing standards that require that all systems must have privacy impact assessments and system security plans approved before operational use of the system;
c. Provides education of all persons utilizing medical data;
d. Establishing security standards and requirements for data and data systems;
e. Establishing security, authentication, and access standards and requirements for data systems;
f. Establishing accreditation standards and requirements for systems;
g. Establishing data classification standards and requirements for data protection based on such classification;
h. Establishing appropriate sanctions for non-compliance of requirements, or altering, falsifying, or concealing any records kept by the government;
i. Performs periodic audits and reviews of national systems to ensure compliance;
j. Establishes specific ownership, confidentiality, and usage standards for all data; whether collected, matched, aggregated, or utilized for research. As a minimum, that patient data are owned by the patient; aggregated or otherwise derived data belongs to the government; the government has rights to use patient data;
k. Establishing standards for use of citizen data by government officials;
l. Establishes 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;
m. Establishes standards and requirements relating to flow of personal information; including written plans and disclosures about the data use; and 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 shall include a description of the personal 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 in regards 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.
n. Establishing a mechanism for release of patient records to the individual to whom the record pertains and requires notification of any release of data to that individual;
o. Requires that all data system users sign an ‘Acceptable Use Agreement’;
p. Establishing data disposal standards and requirements;
q. Develops 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;
r. Encourages the sharing and implementation of best practices in informational privacy management across the federal government and promote the exercise of self-correction, prevention of data loss or misuse, and risk mitigation;
s. Encourage and foster a process of continuous learning and improvement about the codes of ethics that professionals engaged in privacy management need to set for their practice;
t. Establishes that the Office of Information Security can receive complaints concerning the full array of rights and protections under the Privacy Act, including complaints of inappropriate collection, use or disclosure, failure to maintain up to date and accurate data, improper retention or disposal, and complaints relating to 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 institutions response to the complainant. If the response from the institution is not satisfactory, 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; however, they are precluded from providing guidance on standards, requirements, or what constitutes inappropriate collection, use, or disclosure of personal information.
a. Provides a resource for investigating breaches of confidential information whether accidental or intentional and provides recommendations to those responsible on how to better secure this information in the future.

v. Provides for the ability to levy fines commensurate with the extent of the damages.

9 Leveraging Existing National Government and Other Infrastructure

9.1 National Communications and Computer Infrastructure

The IT system containing and/or coordinating the national patient registry will generally require data to flow between a national host and regional hosts (servers), and possibly additional hosts. For this data to flow throughout various locations within the country, a communications infrastructure will need to exist that will robustly support the movement of data. The infrastructure can consist of public and private networks, but will have to have sufficient capacity to support significant daily flows of information from host to host in order to achieve data synchronization across a number of hosts.

In order to keep the amount of data that is required to synchronize dozens to even hundreds of systems which contain the patient index, a software architect who is experienced in Value Added Network (VAN) design, deployment and use should be consulted. This architect will be able to evaluate needs of the patient registry within the particular environment of the country and generate projected communications requirements necessary to keep all systems up to date. These requirements will need to project communications requirements and recommendations depending on a number of rates of enrollment, and ongoing information flows of enrollees.

There will be a minimum of a three-tiered environment necessary in each country; the top tier consisting of one or more national systems configured as a stand alone or clustered systems; the second tier consisting of regional hosts; and the third tier, the clinic or hospital endpoint of the data. Depending on the size of the country, it may be appropriate to have sub-regional (such as district) level hosts in between the regions and the endpoints. The number of tiers needed within the country will depend on the size of the country, the robustness of the communications systems, and the volume of the patient registry transactions.

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

- Air and power conditioning to ensure that the computers can operate at least at a 95% availability (or what ever availability is desired);
- Physical access control and accountability to the locations where the computer is housed;
- Application access control;
- System logging, and active or periodic log analysis;
- Robust national-level pass-fail system and application monitoring;
- Support mechanisms in order to quickly resolve computer system and communications failures and errors.

It also should be noted that server hardware used in modern data centers may not be the appropriate choice for many countries. Such equipment demands a consistent level of power and air conditioning that may far exceed what can cost-effectively be provided. Today hardware that is intended for use in embedded applications (fan-less, low-power) may be a better choice for national level deployments. Careful considerations need to be made to avoid expensive and often unsustainable hardware.
9.2 Sociological Realities that May Impact Adoption

9.2.1 Stigma Avoidance

In order to avoid possible social stigmas, careful consideration should be given to how cards and identifiers are deployed. For example, if the number is issued via a printed card, it is generally better to do this for all infants at birth or all individuals in a community. Deploying cards according to a specific disease or condition may increase stigma.

When implementing the NHID across the entire facility of services, the following is a recommended priority list for staging deployment:

- Central (regional) hospital – Outpatient reception, inpatient reception, patient records (to retrofit existing patients), and then other reception points;
- Satellite clinics of the central hospital, if any;
- Other, smaller hospitals and general healthcare facilities;
- Specialized clinics

10 Considerations for Using Modern Identification Card Systems

10.1 Deciding the Type of National Health Identification Cards

The National Health Identification 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. If the card is a ‘smartcard’ and therefore contains a memory chip, it can also be used to further the identification of the patient by storing biometric identifiers, and in addition it could store patient information.

Considerations for deciding on the type of identification card are:

- All cards will require a properly trained person that can operate the equipment needed to generate a card. This equipment will consist of a computer, card printer, a camera, and possibly a laminating machine.
- Plastic or plastic laminated paper cards are relatively easy to make and inexpensive to generate.
- Cards containing electronic circuits (smartcards) are considerably more expensive, and prices vary depending on the amount of memory on the card. Cards typically cost $0.75 to $3.00 USD, depending on the volume and type, including printing. In addition, a smartcard infrastructure requires card management systems; card issuing machines; many card readers; and significant training of card administrators.
- Smartcards can provide significant additional value in identifying the patient by providing biometric identification data; and by providing core patient data, latest diagnostic and treatment, and other relevant data.
- Smartcards contain electronic components which can be damaged by bending, twisting, or perforating the card and as such are less sturdy than plastic or plastic laminated cards.

10.2 Biometrics

Biometrics can be helpful in identifying an individual based on one or more physical trait. The most common physical traits utilized in recognizing an individual include fingerprints, face recognition, photos and iris scans. While fingerprint scanners are relatively inexpensive and don’t require extensive training to use, it is important that the exact models that are being considered are tested in actual use conditions, with the populations that will be using them. Since the scanners depend on consistent physical...
characteristics of the person being scanned, 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 are $75 - $150 USD. These scanners utilize a simple optical method of recognizing the ridges in the fingerprints. Forensic quality optical scanning fingerprint readers are $400-$700 USD. These scanners are a higher-quality version of the typical optical scanners, and have much higher resolution. All optical scanners are affected by: skin dryness; how much water consumption has occurred; low temperatures, and the condition of the skin. They typically have a 65% - 85% success rate. They will not function on children less than 5 years old, people who wear their skin down when performing manual labor, or the elderly.

Fingerprint scanners that rely on recognition of vein characteristics in addition to the skin ridges are more accurate and some can be used in more harsh conditions. The inexpensive units cost $125 - $300 USD depending on the quality, and the ones built to be used in more harsh conditions with moderate amounts of water and dust, are $600 - $1000 USD. These readers are currently being used within systems where more accuracy is needed, including bank ATMs and other banking functions to authenticate users.

Iris and face scanners are in the range of $900 - $4000USD depending on their accuracy, and typically require a software development kit which costs $500 - $1500 US. In addition, they require a more 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 scan, they are less suitable for use in countries that do not have robust, high-speed networking to all facilities.

Whatever technology is chosen, the biometric readers will need to be available at all patient registration and presentation points, which could possibly number in the thousands. In addition, training will be required, but this can typically be accomplished at the facility level.

10.3 An Ideal National Health Identifier

A simple user-friendly NHID that is suitable for use by both humans and computers constitutes an ideal choice for the NHID. There have been significant amounts of work completed by international work groups to think through the design of a health identification number, and this work has yielded excellent guidelines for determining such a number. There exists now a well-developed concept called the Sample Universal Health Identification (UHID) Number. The characteristics of this number have designed by ASTM in the form of Conceptual Characteristics. These characteristics also apply to an NHID, and include: 1) functional characteristics; 2) linkage of lifelong health record; 3) patient confidentiality and access security; 4) compatibility with standards and technology; 5) design characteristics; and 6) reduction of cost and enhanced health status.

1. Functional Characteristics:
   a) Accessible: Access is dependent upon the establishment of a network infrastructure, the trusted authority, and policies and procedures that support the system.
   b) Assignable: Assignment of the Sample UHID or Encrypted Universal Health Identifier (EUHID), 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 will also depend on the mechanism to request a Sample UHID.
   c) Identifiable: This will depend on the identification information that the trusted authority links to the Sample UHID.
   d) Verifiable: The Sample UHID includes a six (6)-digit check digit for verification.
   e) Able to be merged: The internal data structure of the Sample UHID does not directly support merging duplicate or redundant identifiers. They can be linked at the trusted authority.
f) Able to be split: There is no inherent support for splitting the Sample UHID. New IDs can be issued for future use. Splitting for retroactive information must be handled by the trusted authority.

2. Linkage of Lifelong Health Record
   a) Able to be linked: The Sample UHID has the ability to function as a data element and support the linkage of health records in both manual and automated environment.
   b) Able to be mapped: With the use of appropriate database system and software, the Sample UHID can be used to map currently existing healthcare identifiers.

3. Patient Confidentiality and Access Security
   a) Content Free: The Sample UHID is free of information about the individual.
   b) Controllable: This depends on the policies and methods that will be adopted by the trusted authority.
   c) Healthcare Focused: The Sample UHID is recommended solely for the purpose of healthcare application.
   d) Secure: The Sample UHID includes a EUHID, which offers a mechanism for secure operation through the use of encryption and decryption processes. These capabilities depend on the policies and procedures that will be implemented by the trusted authority.
   e) Dis-identifiable: EUHID supports multiple encryption schemes offering multiple EUHIDs to prevent revealing the identification of the individual.
   f) Public: The EUHID'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.

4. Compatibility with Standards and Technology
   a) Based on Industry Standards: The Sample UHID is not based on existing industry standards. It is based on ASTM's Standard Guide for Properties of a Universal Healthcare Identifier (UHID).
   b) Deployable: The Sample UHID is capable of implementation in a variety of technologies such as scanners, bar code readers, or other uses.
   c) Usable: The Sample UHID is capable of implementation in a variety of technologies such as scanners, bar code readers, etc. The 28 digit identifier will present difficulty for manual computation and transcription. It may be a time-consuming process and subject to human errors.

5. Design Characteristics
   a) The ASTM guide and the proposed Sample UHID do not address the implementation issues and infrastructure requirements.
   b) Unique: The trusted authority will be responsible for the uniqueness of the Sample UHID.
   c) Repository-based: The Sample UHID can be stored in a repository.
   d) Atomic: The Sample UHID consists of a sixteen- (16) digit sequential identifier, a one- (1) character delimiter, a six- (6) digit check digit and a six- (6) digit encryption scheme. It can function as a single compound data element.
   e) Concise: The Sample UHID is not concise. It is a 29-character identifier.
   f) Unambiguous: The Sample UHID is unambiguous. It uses numeric characters and a period as a delimiter.
   g) Permanent: The Sample UHID has sufficient capacity to prevent reuse of identifiers.
   h) Centrally governed: This policy issue is not addressed. The Sample UHID requires central administration and is dependent on the establishment and functions of a trusted authority.
   i) Networked: The Sample UHID can be operated on a computer network. It requires establishment of the necessary network and technology infrastructure.
   j) Longevity: The Sample UHID can support patient identification for a foreseeable future.
   k) Retroactive: Has the capacity for retroactive assignment of the Sample UHID to every person in the country.
   l) Universal: Can support patient identification for the entire world population.
   m) Incremental Implementation: The Sample UHID 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 UHID and existing patient identifiers can coexist during the time of transition.

6. Cost and Enhanced Health Status

   a) The Sample UHID has the potential to support the functions of a NHID. The establishment of both the administrative and technology infrastructures; the creation of a Trusted Authority; the design and development of computer software, hardware, and communication networks; and the implementation security measures will require a substantial investment of resources, time, and effort.

10.4 Calculating the Size of the National Health Identifier

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

The things that must be considered to calculate the minimum required length of a NHID include:
- Base population
- Average age of a generation when bearing children
- Average lifespan
- Amount of years past death that a number should be unique (the quiescent period).
- Population growth averages.

As an example, assume that we will be using the NHID in a country where the population is estimated to grow 20% in years 2000-2050, and assume that same growth pattern for the following 50 years (totaling 40%). The current population is 100 million, so the projected population in 100 years would be 142 million. With an average age of birth mothers being 22 years old and an average lifespan of 66 years, to identify all persons from birth to death and include a 200 year after death quiescent period before the number would be re-issued; we would need a patient identifier that would have a capacity for approximately 660 million. This nine- (9) digit number would easily fit into a twelve- (12) digit NHID. This would include the allowance for a final digit being a check digit, and allow for two orders of magnitude including extra space to ensure that the number is sufficient in the future.

The recommended check digit calculation would utilize the UPC check digit formula. See Appendix C for further details.

11 Defining, Implementing, and Sustaining National Standards

11.1 National Data Standards

National data and meta-data standards provide a common understanding of data elements for a national “core data set” 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. While 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:

1) All in one variable. Even that can be confusing since the name could be typed in as ‘First name, Last name’ or ‘Last name, First name’.
2) Separating out First, Middle, and Last names.
3) Prefix, First, Middle, Last, and Suffix.
4) Middle can be one name, several names, or just an initial.
5) Also consider Also Known As (AKA, or alias) names

Another example is for laboratory tests, are the tests coded the same in all systems? And are the results also stored in exactly the same way?

As these examples illustrate, it is very difficult to share data from one system to another when each data element or variable in the system can be stored or coded in a different fashion. Since all data has a meaning in which it is being used, if we can agree on a single term and its associated meaning and format from the beginning, then we will be able to share the data with other systems. Terms and data agreements also set the basis for software development by standardizing items and item descriptions. This can also simplify the coding of existing reporting tools and search engines that can pose queries against collections of data.

The national minimum data set should consider data items for registration and scheduling, monitoring and improving practice management, and describing clinical encounters and clinical care. While various countries have defined what is included in their minimum data set, these usually include:

- National Health Identifier;
- Patient name and address or other locator information;
- Other patient demographics;
- Family identification links;
- Encounter date;
- Care facility;
- Medical history;
- Physical examination;
- Diagnostic test and results;
- Treatment data including drugs, procedures, physical therapy, and counseling;
- Visit diagnosis;
- Medication adherence;
- Referral to other facilities;
- Adverse events;
- Death date and cause.

Data should be numeric or coded using standard definitions and no or minimal free text. To improve accuracy, efficiency, and availability, data should initially be entered electronically at the source, and that source system should be considered the system of record. All other systems utilizing that data should receive it electronically from the source system, thus preserving the integrity of the source data.

Data elements must be sufficiently detailed to support clinical algorithms, guidelines, and aggregation into broader categories for consumption by higher-level users. The proposed minimum data set will evolve over time as funding becomes available, care protocols change, and additional tests and treatments become available.

To enable a common data dictionary at a national level, there must be a national medical data standards authority, which has been empowered with the appropriate governance authority and the mandate to proceed in building a centrally-managed database of standardized terms, data, and values.
11.2 Action Items for the National Medical Data Standards Authority

The data standards authority should commence regularly scheduled meetings with the appropriate management for the following disciplines:

- Patient care providers, including doctors, nurses, counselors, and other clinical staff;
- Laboratory workers;
- National government healthcare oversight and reporting;
- Information technology;
- Non-Governmental Organization (NGO) workers and advisers in the healthcare arena;
- Other staff as appropriate.

The data standards authority should first determine the appropriate scope of specific medical disciplines in which terms should be standardized. Once agreement of the initial scope is agreed upon, this group should examine what internationally accepted standards exist for the various disciplines.

Universal code systems are now available for subject matter such as units of measure (ISO+ - International Organization for Standardization), laboratory observations (Logical Observation Identifiers Names and Codes - LOINC), common clinical measurements (LOINC), drug entities (NDC), device classifications (Universal Medical Device Nomenclature System - UMDNS), organism names, topology, symptoms and pathology (Systematized Nomenclature of Medicine--Clinical Terms ; SNOMED), International Union of Pure and Applied Chemistry - IUPAC, outcomes variables (Health Outcomes Improvement - HOI), and location codes (ISO 3166). Some of these coding systems are available at no or reduced cost for lower and middle-income countries.

Upon completion of an initial scope, an examination of the status of international standards, and identification of which items will need to be defined on a national level, a report will be needed enumerating the current status of standards what standards or subset should be generated. This report should include what, if any, changes would be needed to incorporate those standards consistently on a national basis.

Upon issuing the report and having discussions about findings, the group should generate a project scope document that enumerates which items will be prioritized and presents an overview as to the schedule of organizing the project and enumerating the resources that will be engaged to perform the needed tasks.

The group should then generate a communications plan, and schedule initial meetings to engage the appropriate stakeholders and subject matter experts to build consensus on the national database metadictionary.

The results of the national medical data standards authority should be published periodically, and it is recommended that they be available at any time via a web portal.

11.3 Data Exchange/Sharing Standards

Once having engaged both national data standards and national data sharing standards, it is possible for data to feed from one data system to another in a manner that benefits many. Internationally agreed upon standards that govern movement of data amongst systems exist. With these standards, many problems can be solved and a first-stage medical record system created from the extensive medical data that already exists in systems such as laboratory, pharmacy, scheduling, and case management systems.

Standard mechanisms for communicating securely over networks in a secure fashion exist, as do standards for delivering structured medical record content like patient registry records, orders, test results, and standard identifiers for coding most, of the concepts we want to report in the fields of such structured records.
The communication standards of choice are the internet standards, and include the base internet protocol for sending packets of information, the secure sockets layer (SSL) for encrypting transmitted information, certificates for verifying the identity of the communicant, secure File Transfer Protocol (FTP), and secure (email).

There are many well supported and easily used standard Internet protocols that should be considered to exchange health data. Available security tools are more than adequate for the additional threats that are encountered when utilizing the public Internet.

There are also several health specific data exchange standards that should be considered:

1. HL7 is the message standard of choice for communicating clinical information such as diagnostic results, notes, referrals, scheduling information, nursing notes, problems, clinical trials data, master file records, and more. It is used by more than 2,000 hospitals, by the US Centers for Disease Control and Prevention (CDC) for immunization, communicable disease, and emergency visit information, as well as by most large referral laboratories. It is also widely used in Canada, Australia, Japan, and in many countries in Europe.

2. HL7/ASTM (American Society for Testing and Materials) provides the structure—like a set of database records—for interchanging patient information between source systems like laboratory, pharmacy systems, patient management systems, and other medical record systems.

3. Digital imaging and Communications in Medicine (DICOM) is the standard of choice for transmitting diagnostic images. It is supported by all imaging vendors, and is working closely with HL7.

These health data exchange standards do not typically specify the choice of codes for many fields. They do provide a mechanism for identifying the code system for every transmitted code. This pluralistic strategy was the only alternative in the past because universal code systems did not exist for important topics such as laboratory tests and clinical measurements; so institutions used their own local codes. The universal coding system mentioned earlier – ISO+ (units of measure), LOINC (laboratory observations), NDC (National Drug Code), UMDNS (device classifications), SNOMED and IUPAC (organization names, topology, symptoms and pathology), HOI (outcomes variables) are mostly available without cost. So, for at least some source systems, we have all of the pieces needed for creating Electronic Medical Records (EMRs) utilizing data from multiple independent sources, inside and outside of a health care organization. Note that some of these standards have costs associated with their use.

12 Aggregate Reporting

Generating accurate aggregate data is difficult at best when there is no universal patient identifier. In the absence of NHID, attempts are made to use population surveys to provide adjustment factors. While this can be mainly successful, it does not provide a high degree of accuracy, nor does it contain detailed justification for the aggregated numbers in order to provide high levels of accuracy.

To obtain high quality aggregate data, we must rely on:
• Well-sourced, complete, and non-redundant patient counts
• Meta-data standards
• Data exchange standards

Fortunately, most of these items are a natural outcome of implementation of a NHID, medical data standards, and national data sharing standards.
13 Data Quality Integration and Auditing

To overcome the issues of inconsistent data quality and data duplication, regular data audits should be performed. Some of this effort should be directed at integrating data from multiple and heterogeneous source systems. Through the process of data quality validations, exceptional data can be identified and corrected.

Data integration is a continuous process, and it is essential to impose regular data quality audits, reporting, analysis, and correction. In addition, the following country-wide activities should be incorporated and improved over time:

- A strong data governance team to guide source data owners and data stewards for better data quality, and to assure timely and coordinated data collection;
- Domain and subject matter experts to lead the data quality assurance team;
- A data quality enforcement policy;
- Data quality and standardization rules;
- Data quality metrics to measure the trust factor.

14 Conclusion

The successful implementation of a NHID is a complex process. It is important to understand the series of related efforts that countries may be required to address the process successfully. Considering this, the following should be considered:

- The amount of effort required to complete each step
- The priority of the various items
- Not all things are possible to be completed at the same time
- Items should be grouped together logically, and progression of work on the items in each group should be done concurrently when possible in order to speed up the process.
- Every implementation of a NHID will probably take two to five years to be fully integrated across the organizations that will use it. The goals set of for example, 90% of the eligible population having been issued a number, will take at least five years, and probably take 10-20 years depending on the level of resources applied and aggressiveness of the implementation plan.

The top ten challenges when implementing a NHID are:

1) Institutionalizing standards-based locations and other national reference data. Many data silos are already in place in most countries, typically at the facility level, but possibly at the section level within a facility. For a national identifier to be useful in collating disparate health records, all current location information must be identified and numbered in a standardized way at the national level. This location information must be detailed in order to support the lowest level of locations used within the country, possibly down to the clinic level. To implement this, there must be standardized location information for the highest tier of location information (for example, regions), the second tier (for example, districts), and all other logical divisions in the country. There is an ISO standard for country numbering and first tier numbering (regions). All other tiers must be numbered in a nationally-standardized manner, which would probably include city, township, or similar location possibly including sub-township location, all the way down to the point of service provision. Geocoding location information is encouraged in order to specifically locate a patient’s home address, as well as servicing facilities. The lowest level of division is the

6 The level to which data should be trusted to make critical diagnostic or operational decisions.
service providers, who also must be uniquely identified for this information to be collated together and uniquely identify a health facility.

2) It is not possible to have good data in a national data warehouse without a strong patient identification system in place. To support aggregated information, strong de-duplication methods must be utilized, which require a strong NHID implemented within a national health registry to tie together the numerous health IDs in existence.

3) Consider the scope of the effort – is it practical to develop a “universal identifier” for all citizens for all governmental purposes? Should an identifier be for health use only? What are the practical political issues that surround multiple ministries having buy-in and their agreement about a “universal identifier”? Should there even be a universal identifier, which if used in the wrong way could cause a financial loss or other significant problem to an individual?

4) The identification number 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 not be re-used within at least 200 years of a person’s death. If the identification number will also be used for real estate transactions or other similar long term use, it will need to be sufficient in size to be utilized for more than 500 years.

5) The number should be issued at birth, tied to all health records from laboratory, multiple clinics, and other sources while alive, and eventually tied to death certificates. If medical records are to be standardized about pregnancies and going forward from that point, they will need to be issued at the time of the first visit to any health provider working with pregnant women. Also, the development of standardized datasets for birth, clinical encounters, demographics, death, possibly pregnancy, and other items should be considered as part of the process.

6) The identifier should never include data. Data will change over time; and it is not a matter if it will change, but only a question of when it will change. At the point in time when the data changes, which might include location information, all numbers that have been issued will be invalid and at least require translation or re-issuance. Also, if data are embedded in the number, at least some amount of privacy information will be publically displayed.

7) The number should include a check digit at the end to minimize entry errors. There are standardized, simple algorithms such as Universal Product Code (UPC) that can be used.

8) The NHID, if containing no data and if dispensed sequentially 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 risk.

9) For indicators ever to be accurate, the national identifier will have to be utilized to validate the aggregate counts. For example, if a patient moves from one pharmacy to another, the notion that the patient was served up to a certain point by one pharmacy and then by another when they relocated must be understood in order to not count as multiple patients when counting the annual number of patients served.

10) A NHID will need to be accompanied by a national health registry in order to properly collate the different sources of patient data, which are identified individually at the facility level. This registry will have basic demographic data, and a variable number of identifiers used by location to identify the patient. A well-documented plan, including estimates for the amount of work required to implement a health registry, should be completed up front to fully comprehend what is required for a successful implementation.
Appendix A: Unique Identifier Systems – Key Concepts

Conceptual characteristics of unique identifier systems

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 healthcare identifiers. For an in-depth treatment of each conceptual characteristic, please see ASTM 1714.3. The degree to which a given unique identifier satisfies these characteristics will influence the degree to which the identifier can support the healthcare functions described in section II.A. Related conceptual characteristics can be further divided into six subgroups1, which include:

1. Functional characteristics describe global system-level capabilities that are enabled by the identifier. These characteristics include being accessible, assignable, identifiable, verifiable, mergeable and splittable. 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 ID and consists of a single digit computed from the other digits in the identifier.

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

3. Confidentiality and security characteristics address the ability of an identifier to protect and preserve patient privacy. These characteristics include being content free, controllable, healthcare focused, secure, de-identifiable and public. For example, a healthcare focused unique identifier is used exclusively in the context of healthcare and does not extend to other enterprises. A content free identifier contains no personally identifying information, 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, so for example if patient’s year of birth is contained in the identifier, systems may begin using that information for system functions such as patient registration, age determination, etc. If the identifier format alters or removes year of birth information, the functionality ceases to work. It is important to note that the principle of being content free stands in direct contrast to biometric identifiers.

4. 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 implementable using different technologies such as smart cards, barcode readers and paper. While many existing standards-based health care transactions accommodate unique identifiers, one should 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. (http://www.hl7.org) Identifier schemes that use more than 20 characters may be incompatible HL7 version 2.

5. 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 minimize recording errors, the time required for use, and storage required. Contrast this with a long
lasting identifier, which must be sufficiently long to accommodate substantial information content that will scale with populations that increase in size over time.

6. 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 number.** Creating sequentially increasing integers, with or without leading zeros (0001, 0002, 0003, …) 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 for creating identifiers similar to serial numbers involving alphanumeric codes and special formatting exist. The U.S. 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 (GUID/UUID).** 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 (though non-zero). A current commonly used GUID scheme produces 3.4×1038 possible values, and the methodology allows unique identifiers to be created locally on any computer (a distributed assignment system). An example of such a GUID is: `{3F2504E0-4F89-11D3-9A0C-0305E82C3301}`. While GUID’s ensure a high degree of uniqueness, they do not explicitly have a check digit for self verification and also contain many digits. The GUID’s notable length and lack of check-digit limit its usefulness for manual data entry applications.

- **Quasi-unique personal identifying elements.** While not guaranteed to be explicitly unique, a combination of nearly constant personal demographic elements is commonly used to establish unique identity. These elements can include given name, surname, sex, birth date, mothers first name, mother’s maiden name, birth father’s first name, birth location (such as country at birth, or country and city at birth) and birth order (1, 2, 3, etc.), among others. 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 unique. While the identifiers are guaranteed unique, this system may assign multiple identifiers to a single patient. This can occur when patients receive care at different sites without their previously assigned identifier. One potential security limitation of this approach is that the site type (such as ART clinic) may be embedded or implied in the site identifier. If such codes are publicly available, the identifier might not be fully content-free.

- **Biometrics.** Biometric identifiers, including voice patterns, fingerprints, iris patterns, facial shapes and vein patterns, are another 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. And 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 (disappear) with time, and retinal patterns change for very young or old patients, and also can change in patients with conditions that affect the eye, such as diabetes. Additionally, privacy concerns remain 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 and by their nature, biometrics are effectively non-revocable. Further, resource
constrained settings may not be able to feasibly accommodate the technology required to manage biometric identifiers, which may limit their utility in developing settings.

- **Identifiers with additional functionality.** In addition to ensuring uniqueness, health care identifiers may convey additional features such as privacy and security. The ASTM 1714 document describes a process to include security information in the identifier to indicate whether a particular identifier is intended for fully-identified use (e.g., in patient care) or as a private limited-use token (e.g. for population level reporting or research). While incorporating additional features into the identifier may be desirable from a system implementation perspective, it’s important to note that the identifier then compromises the principle of being content free.

**Conclusion**

Domains such as law, health policy and ethics provide inputs that inform the strategy for deploying unique health care services identifiers. As summarized in this appendix, technical and process inputs also inform the unique healthcare 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. Further, the preferred system requirements and existing operational constraints that must be accommodated should be identified to ensure broadest use of the system.

In general, it is never a good idea to embed data in a patient identifier, since the identifier will be in use for hundreds of years, and data such as region or district codes can change over time. When the data changes, it probably invalidates the existing patient numbers that are based on the old codes, and therefore causing a significant amount of effort to remediate the changes.
Appendix B: Checksum Calculation for UPC checksum

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.) together and multiply by three.
2. Add the digits (up to but not including the check digit) in the even-numbered positions (second, fourth, sixth, etc.) to the result.
3. Take the remainder of the result 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 UPC-A barcode for a box of tissues is "036000241457". The last digit is the check digit "7", and if the other numbers are correct then the check digit calculation must produce 7.

1. We add the odd number digits: 0+6+0+2+1+5 = 14
2. Multiply the result by 3: 14 × 3 = 42
3. We add the even number digits: 3+0+0+4+4 = 11
4. We add the two results together: 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. Therefore, the check digit value is 7.
Appendix C: Example of Interactions between a Health Facility and the National Health Identifier System

Patient received in OPD other dept. →
Health passport examined to see if new patient

Is there a national ID on passport?

Yes →
Done

Facility software stores response, prints patient ID label for passport or other papers

No →
Connect to Regional Health ID system, pass message to request verification or new health ID

Is it a new Patient?

Yes →
Issue a new national health ID along with any known local facility IDs

No →
Issue existing patient ID and any known facility IDs
Appendix D: Example of Local Health Provider Systems Interactions with the National Health Identifier System

1. Health facility software reads healthID.cfg file at start, which contains list of 1 or more (3 suggested) IP addresses for National Health ID system servers.

2. Health facility sends out “Are You Alive” request every minute to all servers in its list.

3. National Health ID systems respond back with Yes.

4. If no response after 30 seconds, socket is closed and request is terminated.

5. Health facility updates status table with response time in seconds. No response = 999.

6. Lowest number in status table (lower than 999) is used to request National Health ID.

7. All Patient ID service requests are encrypted using AES encryption.

The National Healthcare ID Value Added Network
Appendix F: Implementation Models and Resource Planning

The value of a national health identifier (NHID) system depends on the quality and availability of the data used to identify the patients. In order for the system to perform well, the applications must be clearly understandable and made readily available to the applicant, processed correctly and in a timely manner. And finally, the resulting identification card be delivered to the patient and the information verified in a punctual time frame.

Implementing a national health identifier (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:

1) The specific implementation model which decided upon.
2) Population and population distribution.
3) Expected monthly enrollment rate.
4) Number of concurrent locations providing enrollment services.
5) Whether publically available enrollment places can utilize existing building space.
6) Availability and costs of staffing the customer contact points.
7) Geographic distances and population distributions.
8) Distribution of the medical ID registry system.
9) Which types of identity verification are going to be utilized, such as photos and fingerprints.
10) Whether data and voice communication systems needed to support the effort are adequate to support this task.
11) Whether the scope of the project includes development of national data standards, and national data sharing standards, at least for identification data elements.
12) The quality and availability of public roads and transportation systems.

Much of the information that follows is technical in nature because the costs are directly driven by the costs of implementing certain technological models. Therefore, it is strongly encouraged for a well organized assessment which yields a clear path to achieve the expected outcome. This evaluation should be an interactive process which includes both information technology professionals with system architectural experience and management. One of the most successful ways of achieving a good outcome with well understood costs is to commence a management stakeholder group which identifies appropriate technical resources and forms a technical working group (TWG) of the technical resources. A set of goals should be identified in order to empower the TWG, and an acceptable calendar set for those goals. By engaging the proper resources, 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.

14.1.1 Implementation Models

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

14.1.1.1 Heavily centralized Model

Characteristics of a heavily centralized model include:

- This model usually minimizes costs since it can probably utilize enrollment locations which are already providing registration services, and possibly are already issuing permits or licenses.
- 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, and application for a drivers or marriage license.
- This model lends itself to centralized administration and issuance of cards.
• Annual costs are typically lower than other models due to the limited number of people applying on an annual basis, and utilization of existing infrastructure to receive applications. Of course, the costs of processing applications, and administrative costs of the national registry are nearly the same cost per enrollee for all models.

14.1.1.2 Semi-distributed Model

Characteristics of a semi-distributed model include:
• Costs are typically somewhat more expensive than a heavily centralized model but less than 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 the events of the heavily centralized model such as birth, reaching a certain age, hospitalization, joining the military, and application for a drivers or marriage license.
• Typically, the application points have card printers and can also issue the cards upon administrative approval.

14.1.1.3 Highly distributed model

Characteristics of a highly distributed model include:
• A much more complex rollout schedule, which means a number of teams performing site readiness to meet an aggressive schedule such as 1-2 years.
• Higher costs due to the higher number of application and issuance points, the number of teams executing the site startups.
• If an aggressive schedule is desired there will be a higher number of rollout teams required, as well as higher administrative costs to process the higher 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 enrollment efforts, and a medical event occurrence such as hospitalization or clinic visit.
• It requires a carefully thought out and available communication system, with robust connectivity to at least the district level, and probably the larger site level.
• There will be higher rollout costs due to larger rollout team, which includes providing equipment, hiring staff, and training on the identification system.
• Unless there is a uniform national language, translation will need to occur at least for the application forms, and possibly the data entry screens in order to fully support multiple languages in use.
• 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.

14.1.2 Identifier Methodologies and Systems

Depending on the personal identifiers that are utilized, suitable verification of the information must be done during the application approval or denial process, and during delivery of the identification card. Personal descriptors might include a picture, name, height, eye color, sex, and age. It is important to keep in mind that all data that is stored in the NHID registry and contained on each identification card must be accurate and verified to the degree that it can reasonably be done. Consistently accurate and timely data will only occur through 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.

14.1.2.1 Identification Cards

ISO 7810 is an international standard for the physical characteristics of identification cards. Specifically, 7810-ID-1 specifies that the cards are 85.60 mm x 53.98 mm. This is the most common size of bank, credit, and debit cards, as well as driving licenses.
ISO 7811 is an international standard for recording of printed and magnetic data on the identification card. 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 (called smartcard) and electrical connections for the chip.

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

### 14.1.2.2 Biometrics

The most common physical traits utilized in recognizing an individual include height, sex, fingerprints, face recognition (photos), and iris scans.

Whatever technology is chosen, the biometric readers will 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 successfully utilize this technology, training will be required. The training may typically be accomplished at the facility level in order to minimize the staff disruptions, or be more centrally located in order to maximize the trainers time.

All prices stated are 2011 prices found from various suppliers on the internet, must be verified and are simply a guide to understand the variances of prices.

**Finger Print Scanners**

Typical inexpensive fingerprint scanners are $75 - $150 USD. These scanners utilize a simpler optical method of recognizing the ridges in the fingerprints. Forensic quality optical scanning fingerprint readers are $400-$700 USD. These scanners are a higher-quality and higher resolution version of the typical optical scanners. All optical scanners are affected by: skin dryness; how much water consumption has occurred; low temperatures, and the condition of the skin. They typically have a 65% - 85% success rate. They will not function on children less than 5 years old, people who wear their skin down when performing manual labor, or the elderly.

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. The inexpensive units cost $125 - $300 USD depending on the quality. The units built to be utilized in harsher (moderate amounts of water and dust) conditions are $600 - $1000 USD. These readers are currently being used within systems where more accuracy is needed, including bank ATMs and other banking functions to authenticate users.

Iris and iris/face scanners are in the range of $900 - $4000USD depending on the accuracy, and typically require a software development kit which costs $500 - $1500 US. In addition, they require a more 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 scan, they are less suitable for use in countries that do not have later model computers plus robust, high-speed networking to all facilities.

### 14.1.2.3 Photographs

Photographs are the most common and by far the most easily understood ways of identifying a person. Utilization of photographs requires four basic components, a computer, a digital camera, photograph management software, a suitable printer, and the identification card itself. When depending on a photograph to aid in the identification of a patient, specific procedures and adequate training must be supplied to persons taking the photograph, persons verifying the data and card, and persons issuing the card to the patient.
Photographs used to identify persons must adequately support both the use on an identification card and the computer monitor. One commonly utilized size for the printed photograph on the identification cards is 358 pixels wide x 441 pixels, printed at 350 dots per inch using the RGB format at 24-bit True Color, prepared using JPEG standardized in ISO DIS 10918-1. The resulting printed photograph is 25.9 mm wide x 32 mm high (1.02 inches wide x 1.26 inches high).

The photograph will be used during the registration administration process, and also could be used at the point of service to identify the person by displaying it on the computer monitor. Therefore, the original photograph should have sufficient adequate resolution to support up to a 100mm x 152mm (4 inches x 6 inches) 24-bit visual at 300 dots per inch.

14.1.2.4 The Photographic Workstation

Photographs will be imaged by the digital camera, and stored on the computer. Properly designed computer software should be utilized to enter the personally identifying information along with digital storage of the photograph.

It is important to physically secure the workstation, camera, and printer both during the day and after hours from theft. These items are typically easily sold and desirable on the black market.

The computer is what performs the actual tasks as defined by the NHID system. It and the 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 one of the newer dual core machines with at least 2 gigabytes of memory and a USB interface for the camera. 2012 prices range from $400 to $800 USD.

14.1.2.5 The Camera

The digital camera or video camera captures a person’s photo and loads it into the software. Digital cameras vary in resolution, speed, simplicity in operation, and ruggedness. Regardless of the model chosen, the ability to quickly retrieve the image from the camera to the computer is an important consideration. 2012 prices range from $75 to $400 USD.

14.1.2.6 The Digital Printer

The digital printer receives all the text, photos, and images from the NHID system software and prints them directly onto a plastic ID card. The specifications can vary on the printer, including speed, whether it supports single or dual side printing, whether it encodes magnetic strips, whether it encodes smart cards, whether it provides lamination, prints in color, and how close it the each edge it will print. 2012 prices range from $1,000 to $8,000 USD.

14.1.3 Communication models

Regardless of the distribution of the enrollment sites and card distribution sites, the overall communication system should be analyzed and a service level decided. The simplest way of developing 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, there will be data from the site performing the enrollment including the enrollment form, and all biometric information such as a photograph that 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 which will need to communicate each other, and decide an acceptable amount of time in hours that it is acceptable that this connection could be unavailable (which will cause a service outage) at a worse case.

The next step is to estimate the amount of data per minute that will need to utilize this communication link. To estimate the amount of data, the types of conversations that occur across the link should be
multiplied by the number of times per hour that it is estimated that the activity will occur. The actual communication link, if it exists, capacity and robustness should be recorded. It is important to carefully determine each portion of a communication link and separately list each. For example, there will be a communications link from a central server to a central network link facility. From this facility, there will be links to other servers. The importance of the first link mentioned is very high, as an outage of this link will affect all connections from all other points to this central server, possibly causing a massive outage.

When recording robustness, consideration should be given for the following:
- What type of physical medium the link uses, such as fiber optic, microwave, or copper wire;
- What the provisioning of the link, such as E1, DS3, broadband, ISDN, 56K, or cellular data;
- Whether the link has redundancy;
- What the estimated percentage of availability is;
- What the maximum outage amount of time is estimated to be per outage;
- Who is providing service to repair outages;
- Are any improvements in the link required to meet the system performance objective;
- Estimated 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 management for agreement on anticipated costs and the noted maximum outage times. Certainly there will be dates in the future when various systems are unavailable, and it management should be in agreement in advance as to what is acceptable. Annual review of these performance and acceptable outage parameters is encouraged, as it will hopefully encourage matching of system performance with what is within an acceptable cost and performance range.

Costs of communications links are highly variable, so a table of the costs of various communications links available by district would be very helpful in aiding the work of estimating costs.

14.1.3.1 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 will be very difficult otherwise for anyone to resolve problems within an acceptable time frame when they are not aware of the problem quickly. There are a number of commercial and public domain software packages which provide link and server monitoring capabilities, and will provide graphs and SMS or Email notifications when certain events occur, such as:
- Server outages
- Server high utilization
- Server disk drive low free space
- Link outage
- Link high utilization
- Link low performance

Typically, these systems will also provide a map for a visual overview of the health of the system, and statistical and numerical reporting of utilization, performance, etc. on demand so that future requirements can be estimated.

Data Compression
Hardware and software data compression can be utilized to improve the ability of a communications link to process transactions. Depending on the type of data, compression of the data can cut the overall volume of the data up to half or less of what it would be without compression. Typically, so called ‘lossless compression’ yields about a 25%-35% rate, so a typical communication link with compression could process about 1.4 times as much data as one 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 and hardware to achieve the compression;
- The costs of maintaining the compression hardware and software;
- Any anticipated additional outages that might be introduced by the additional systems;
- The costs of a higher capacity (if available) communications link to accommodate the data needs without compression.

Often, there is excess capacity in links used in larger cities, and the simplicity of a link without 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 the more complex links should also be considered.

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

**14.1.4 Transaction Processing**

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

At the very minimum, a queue of transactions to be sent, the ones which have been sent, and the responses that have been sent back from the central registry server must be maintained on at 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**, similar to Small Batch Mode, except that they occur in larger batches within a certain number of hours of the beginning of a batch. This could be several times a day, daily, every several days, or even a week of time. The purpose of this is to allow for intermittent or slower communication links such as cellular links, and possibly even sending batches of data via a CDROM, USB memory stick, or some other storage medium.

With any of these links, it is important to consider that a transaction is not of complete until the response has been received at the remote site and processed. So until this occurs, the person entering an NHID application will not know whether it is being evaluated centrally, approved, rejected, or otherwise. Considering this, entering an application at a remote site, batching in less frequent batches of a week and sending it on a USB stick or other similar transportable medium to the registry processing location and receiving a batched result which has occurred for a week could result in not having a response for several weeks.

In most locations, a hybrid model of near real time mode for centrally located enrollment sites, and small or large batch mode will be required. While it is certainly possible, it is unusual for a country to have excellent connectivity to every remote enrollment site. Of course, locating the enrollment sites more centrally or exclusively centrally will result in a much lowered dependency on having adequate country wide communications, at the expense of being less convenient for the persons to present themselves for enrollment, and return to pick up their NHID card.
14.1.4.1 Optimal Transaction Sizes

The importance of engaging a resource with transaction processing expertise has been stated previously. The reason for this is an assessment of estimated data volume from each site, communication links, and transaction capabilities must be undertaken to develop a processing model that will function as expected.

One of the considerations for the evaluation is the amount of data which is associated with each transaction and response. There are software standards for health data electronic data exchange, such as HIPAA in the US, and HL7. The use of data processing standards is encouraged as long as it is reasonable and feasible. With that said, the size, complexity, and efficiency of the transactions in a system like this must fit the capabilities of the communication systems in the country, and can not be so complex to implement that it is not reasonable to do so.

In general, batches of data tend to be more efficient than when the same number of requests are being processed as individual transactions, due to each group of transactions (which could be a single transaction or more) will require identification information as to who the send is and pertinent security envelope information. And each reply contains a similar identification information plus the security envelops, and the responses to the requests. This extra overhead for a few transactions is not important where communications bandwidth is adequate to accommodate whatever might be needed, but could be an important issue for a communications resource constrained setting.

Also, consideration must be given to the trade offs in efficiency verses utilization of international standard protocols. Some protocols, such as HL7 tend to be verbose and therefore will require much larger transaction sets and higher speed communication links than a highly optimized transaction protocol. While it is always considered to be a best practice to utilize accepted standards where possible, it has to be reasonable to do within the circumstances of each country. Therefore, the final step in deciding the transaction protocol that will be used is to estimate the volume to and from each site by building a table of transactions, estimated volume of each type, estimated size of each transaction (utilizing the chosen protocol) and the sum of all of those will calculate the minimal site to site communication bandwidth requirements for those transactions. Since this is such a critical matter, we encourage the engagement of an experienced Value Added Network (VAN) engineer for this portion of the planning.

14.1.4.2 Key Management

The purpose of an authentication key is:

- To be able to identify where a transaction file has come from; and
- To know the data is from a genuine source and has not been modified in any way.

To manage the issuance, expiration, and distribution of keys, a program to do key management must be utilized. There are typically costs per key of $20 - $200 USD, and the program to manage assigning keys comes along with the keys when they are issued from Verisign or another software key system vendor.

14.1.5 National Personnel Requirements

National Coordinator – Reports to the MoH Deputy Director or equivalent. Is the primary management person overall for the project, and is in charge of resources, interfacing with the MoH and other stakeholders.

Public Policy Analyst – Reports to the National Coordinator, and is responsible for analyzing national legislation and public policies which affect the use of public health data and confidentiality and privacy of health data; and providing gap analysis and alternatives analysis as to this project.

Business Analyst – Reports to the Project Manager, and performs scenario evaluation and other situational analysis to assist in the management decision making process. There should be one to six of these persons for the project, depending on the size of the country and the aggressiveness of the project schedule.
Project manager – Reports to the National Coordinator, and provides operational management and communication for the project. There should be one of these for the project, and they should be empowered to make project operational decisions as the lead for the project. This should be a person with excellent organizational and communications skills.

Personnel Receiving NHID applications and Distributing NHID cards – This includes personnel that receive and provide quality control for the NHID applications. The number of these personnel will depend on the number of application points and anticipated volume per location. It is estimated that one person can receive and check the quality of about 50 applications per hour.

Persons performing distribution of the applications – Regardless of whether the applications are received at a location away from where they are processed, then there must be an orderly procedure for the ensure that forms are not lost. They should be bundled into a batch, the batch to be signed for, transported to the processing location, and signed for by an authorized person there.

Processing / approval – Personnel who review the application and run a script on the system to do a cross check for duplicates.

Conflict resolution – A mid to senior level data analyst who has been provided significant training in proper resolution of data conflicts, possible duplication, missing information, etc. There will be at least one of these in a regional or dozens of them in a national office.

Server and communication systems engineers – Reports to the project manager, and provides technical engineering for the hardware and communications systems. In addition, provides technical troubleshooting of operational communications systems and servers.

Software engineers – At least one senior level data system software engineer to evaluate and design software from use cases, and work with the software development staff to ensure proper coding to achieve the goals.

Transaction (VAN) engineer – Person with experience engineering VAN networks and data transactions. This person will provide critical recommendations on communication bandwidth requirements and the core API and formulation of the transaction files.

Software Development Engineers – Persons who are charged with developing the routines, screens, menus, and other software processes and features in order to process data efficiently.

Subject Matter Experts – SMEs are persons who are experts in a particular field, and can provide needed expertise in a particular effort. These are typically the people that have been performing these functions for a number of years.

Testing Personnel – These could be SMEs, but could also just be experienced computer users who have adequate time available to perform in-depth testing.

Data Base Administrators – DBAs are persons with experience installing, configuring, and operating SQL databases.

14.1.6 Public Policy Analysts

A public policy analyst should analyze current public health policies to ensure alignment with the goals stated for this project. The analysis should include evaluating current governance as to whether it contains the specificity to support appropriate confidentiality and safeguarding of health data. The context and scope of these efforts are covered in Section 8 of the document, “Policy and Governance”.

14.1.7 Registry Processing and Administration

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

1) A batch of new applications (could be a batch of one) is received and entered into the registry new application forms.

2) A series of reports are one which evaluates the new applications for completeness and duplication in the system. The output of this report could be a printed only result, or possibly also include output to a batch list of applications which it thinks should be approved, ones that should be
reviewed along with the reason, and those that it thinks should be denied as they appear to be duplicates.

3) Applications which should be approved are then marked as such in the system, or if the previous reports produce a group to be approved, they can be approved in that way.

4) Applications which have been approved but a card has not been produced are printed and sorted, and placed for distribution.

5) Cards are distributed and delivered to the patient. In a highly distributed model, printing probably occurs at the site where the patient is seen, and is simply filed in an appropriate location and presented to the patient the next time they present themselves.

Once an application has been received, it should be checked for completeness. When entering a new application, it should be
New applications processors
Level 1 resolution
Level 2 resolution
Level 3 resolution

14.1.8 Operational Support
Call center junior help desk persons
Call center lead
Senior IT engineer
Junior IT engineers
DBAs

14.1.9 Printing and Distributing Medical ID cards
Typical NHID cards are plastic cards similar to drivers’ license cards. As discussed earlier, they may contain a magnetic strip so the basic identifiers on the card can be electronically read (name, birth date, sex, etc.). If the cards will contain a magnetic strip, there will be increased printing costs because 1) the encoding machines add cost to the printing cost, and 2) the cards will take longer to print and encode, thus extra personnel time in completing the card. An estimate of the costs of a magnetic encoder is XXXXXX and adds 30 seconds to the time needed to prepare the card.

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

• Highly Centralized – Lower cost to print cards, higher cost to distribute cards. Will require cards being delivered to district level locations with persons that are supplying service to (nurses, doctors) or oversight with (community outreach persons) patients; and they will distribute the cards.

• Printing Centralized – Printed on a larger scale / heavier duty printer centrally, and then sending back to the patient point of service to be distributed to the patients. Could be distributed at the next patient appointment or done on an outreach basis.

• Decentralized – Printed at district facilities and distributed to patients at next visit, or during outreach sessions, or via community workers.

• Highly Decentralized – Printed at the points of patient service, and distributed at the next patient visit or via community workers.

The costs of printing the cards is lower when printing centrally, since there would only be several, larger capacity heavy duty printers. The costs of the higher capacity heavier duty printers are 3-4 times the costs of lighter duty slower printers. All printers are rated in expected duty cycles, and printers should be chosen that have at least the capacity to reliably print the number of cards if not more. It is more expensive to have outages and repair bills on a printer than it is to have purchased a unit with the suitable duty cycle initially.

The cost of distribution of the cards is aligned with the number of people that are distributing them. If you are centrally printing cards, then they will need to be sent or shipped back to at least the district level. If community workers or however they are being distributed back at the patients’ residence are going to the patients anyway, there is a small incremental cost of distributing them at the residence. These costs
are mainly time, estimated at 10 minutes per card in order to organize the cards and get them to the person going to the patient, and present the card to the patient with a brief explanation of the card. If the trip to the residence is primarily to give them the NHID card, then the costs will be higher. It will still cost the 10 minutes or so of time per card to organize the cards and get them to the person doing the outreach, and additionally cost the amount of time it takes to go to the patient’s residence. We are assuming that the person doing the outreach is going from patients’ residence to patients’ residence, so there would be no time to add for them returning to the district office to collect another card(s). In order to calculate the time for doing this at the district level, it will be important to talk to the outreach persons and ask them how many persons a day can they visit when do that all day? With that information, you can add 10 minutes, multiply by the number of patients in the district, and you will know how much time is involved. If there are costs involved with the community workers visiting patients, then that should be calculated similarly (cost per visit on average X number of patients per month that are estimated to be seen).

14.1.10 System Architecture

Overall system design, including identifying necessary communications links and setting parameters for those links, resides within the responsibilities of the systems architect. This person will need to work with the VAN engineer in order to understand the characteristics of the transactions occurring. The architect will also need to work with the appropriate persons in order to provide projections of enrollments, enrollment administration and resolution, and system requirements of the centralized, semi-distributed, and fully distributed deployment models.

14.1.11 Equipment requirements

Data entry workstations – there will need to be a data entry terminal for each person putting the application forms into a computer, unless scan-able forms are utilized. If scan-able forms are utilized, there will need to be at least one workstation paired with a forms scanner (OCR scanner) per site. Biometric workstations – there will need to be at least one biometric workstation for each site that is receiving applications. This workstation is where photographs are taken, fingerprints are read, etc., to aid the subsequent identification of a patient. Since this can be a more time intensive process, consideration should be given about having a workstation for each person receiving applications so that long queues of persons do not occur at a single biometric workstation. Central registry servers – There will need to be at least one larger capacity server centrally located, which hosts the NHID patient registry. In addition, there will need to be at least one larger capacity server centrally located which receives transaction files processes them, and generates a response file. 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 (7 days x 24 hours), and also having the capacity to process all requests of the NHID registry. This is quite a significant demand, and in addition will probably mean that any single item that fails, such as a communications link, router, etc., will cause a widespread outage. One way to avoid the single point of failure problem is to utilize regional servers. By doing this, an outage occurring would probably end up just affecting that region, and would entail a simpler situation for troubleshooting and repair. The additional costs of doing this would be the additional servers plus the possible additional bandwidth requirements of each communications link from each of those 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. This means that if there are 5 regions in a country, then there would need to be 5 links with the same high bandwidth to these regional servers as the national server has. Typically these links need to be at least 1 megabit links, and could have requirements of up to 10 megabits in larger countries.
14.1.12 Software Development

The following is a basic list of software development best practices. It is important to note that this document does not attempt to cover this subject in depth, the items being mentioned are there as to be of value in the cost and requirements estimation of the project. The costs and resources required to perform these functions cannot be estimated in advance of performing steps 1, 2, and 3; or at least a high-level version of 1, 2 and 3. More accurate estimates can be performed once a project plan has been assembled by the project manager. The basic steps are:

• Requirements - Gathering and agreeing on requirements is in a successful project. Quality requirements are broken up into two kinds: functional and non-functional. A good way to document functional requirements is using Use Cases.

• Architecture - Choosing the appropriate architecture for your application is critical, since the system will provide a critical new resource nationwide.

• Design – Design of the application should include menus, screens, behavior, and outputs. This step should include SME review of the forms, reports, and functionalities of the system, and the totality of that will constitute a set of system specifications that will ease the job of the code developers. From these items, a project plan is constructed to estimated resources, costs, and a calendar can be developed. In addition, the project charter and communications plan should be developed. It is from this project plan that all resources should be managed, and communications as to progress should be developed.

• Code Development - A best practice for constructing code includes the daily or weekly build and unit test.

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

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

• Performance testing - 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.

• Quality and defects management - It is important to establish quality priorities and release criteria for the project so that a plan is constructed to help the team achieve 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.

• Production System installation and Data migration – Once that the code has been tested and a production date has been 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 SMEs and quality testing personnel.

• Deployment – 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 that might arise.

• System operations and support – Once the system is in production, ongoing support personnel will be required, such as systems administrators, data base administrators, and communications engineers.

Additional information about project management can be found at:
http://en.wikipedia.org/wiki/Project_management
http://www.projectmanagement.com/
There are many other good sources for information on the project management subject.

14.1.13 Training requirements

There will be significant training requirements for the NHID project, including at least the following:
• Application review and acceptance
• Data entry of application
• Registry first level reporting, maintenance, and anomaly review
• Registry second and third tier resolution
• ID issuance
• ID distribution

14.2 Training on Location and 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 providing a Computer Based Training (CBT) course may be the most cost effective way of providing training. This course may be developed by recording video and audio of a teacher who is teaching, or by other methods, or combination of method. If a CBT format is used, we encourage a chapter by chapter test be part of the training video. The results of the data should be stored in a fashion that ensures the integrity of the testing process.

Training at centralized locations requires larger training facilities, and is typically the least expensive way of proving training, if the travel costs involved in bringing the personnel to the location, and housing and feeding them are low. Typically, the travel and per diem costs for all of the persons to be trained far exceeds the costs of sending the trainers to a region or district and doing the training there. Centralized training requires less training resources as the classes can typically be larger than those which are located directly at facilities.

14.2.1 Costing

Estimating the financial requirements for a national health identifier 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 that is processing the applications.
• Office space, computer equipment, communications links, staffing, and staff training for each location where health identification cards will be manufactured.
• 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, along with any consultants who will develop and manage and communicate all project plans.
• A technical working group (TWG) who is evaluating the status of the project, and management committee, who is providing oversight and management decisions for the project.

14.2.2 Summary

Implementation of a National Health Identification System is a complex and time consuming task. This is not to say that it should not be undertaken, but if the project is going to have a successful outcome, there will need a reasonable amount of resources available to complete the task. The easiest way of lowering implementation costs is to have the simplest model which is implemented over a generation of time. In other words, as persons reach a certain age they have an NHID card issued, and they are added into the national registry. The most complex model is a highly distributed model which is aggressively implemented.
Appendix G: Terms and Abbreviations

American Society for Testing and Materials (ASTM) International

Antiretroviral Drugs (ARV)

Clinic

Digital Imaging and Communications in Medicine (DICOM)

Electronic Health Records (EHRs)

Electronic Medical Records (EMR)

Encrypted Universal Health Identifier (EUHID)

File Transfer Protocol (FTP)

Globally (or Universally) Unique Identifier (GUID/UUID). Because a GUID identifier is virtually guaranteed to be unique, this identifier scheme permits identifier assignment in the absence of reliable communication among distributed sites. However the sophisticated code generating method requires a computer. Standard programming languages usually provide a function call to generate this value. Even the compact version of this identifier is not people friendly and is difficult to transcribe efficiently without error. These codes represent good candidates for internal patient, encounter, and site identifiers.

Health Level Seven (HL7)

Health Outcomes Improvement (HOI)

Information Communication and Technology (ICT)

International Union of Pure and Applied Chemistry (IUPAC)

International Organization for Standardization (ISO)

Logical Observation Identifiers Names and Codes (LOINC)

Master Patient Index (MPI)

Multipurpose Internet Mail Extensions (MIME)

National Drug Code (NDC)

National Health Identifiers (NHIDs)

National Patient Identifier (NPI)

Non-Governmental Organization (NGO)

PMI – The Project Management Institute is an international organization which provides testing and certification of project management professionals, and also provides recommended practices and produces books on project management

PMTCT (Preventing Mother-to-Child Transmission of HIV)

Prince2 – A project management best practices set of standards used in the UK and internationally.

Secure Sockets Layer (SSL)

Smart Cards – A card which contains embedded electronic circuitry, and may be used to identify an individual as well as store data about the individual. It requires a smart card reader to read and update the electronic portion of the card,

Systematized Nomenclature of Medicine--Clinical Terms (SNOMED)

Trusted authorities - A trusted authority refers to an entity that performs functions crucial to a unique healthcare identifier system. There may be a single trusted authority assigning all identifiers for a given
system, or there may be multiple trusted authorities operating in a distributed and coordinated fashion. The following non-exhaustive list describes functions a trusted authority may perform:

- Implements agreed upon policies and procedures for granting access to existing identifiers
- Maintains confidentiality and security by providing agreed upon encryption/decryption and other data protection processes
- Responds to requests for assigning new unique identifiers
- Can maintain linkages between multiple identifiers for the same patient
- Ensures uniqueness of assigned identifiers (guarantees no duplicate identifiers assigned)

Tuberculosis (TB)

Universal Healthcare Identifier (UHID).

Universal Medical Device Nomenclature System (UMDNS)

Value Added Network (VAN)

Voluntary Counseling and Testing (VCT)
Appendix H: References and Endnotes


Endnotes


ii Vein Recognition Use Grows is available at http://www.securitymanagement.com/article/vein-recognition-use-grows

iii ASTM E 1714 can be found at http://en.wikipedia.org/wiki/ASTM_E_1714

iv Vein Recognition Use Grows is available at http://www.securitymanagement.com/article/vein-recognition-use-grows