Promoting Interoperability of Health Insurance Information Systems Through a Health Data Dictionary

SERIES INTRODUCTION
By Kate Wilson, Senior Program Officer, PATH

Globally, countries are expanding health coverage to more of their citizens through the development of national health insurance schemes. While the strategies, policies, and technologies used to support these national health insurance schemes are as varied as the countries implementing them, one common challenge is continuously cited—achieving consistent data exchange, which generally relies upon establishing interoperability of existing health insurance information systems.

Why is this data exchange so important? Because the ability for a country to care for its citizens depends upon the ability to identify those citizens, enroll them, treat them when they are ill, and follow up when needed. In turn, providers need to be paid, governments need to be able to track expenditures, and most importantly, the insurance scheme(s) must remain solvent in order to care for all citizens. These "actions" all require data and information exchange, whether paper or electronic based, and all are crucial if the goal of universal coverage is to be achieved.

The ability to perform the above functions is hampered today in many countries by fragmented health insurance and hospital information systems. Existing health information system components in most countries were designed to solve a specific problem but generally weren't designed to communicate with each other. Across insurance schemes and between health care facilities, this lack of a common "language" complicates the exchange of information about patients, diagnoses, payments, and other data needed to provide quality care and facilitate transactions in the health sector. Pharmacies, private providers, community health centers, hospitals, and insurance payers often have their own separate codes, protocols, standards, and technologies that prevent data sharing. This poses a serious challenge, leading to reimbursement delays, increased transaction costs, inefficient use of resources, and the potential for inequitable treatment and fraud.

The rise in the use of computer networks, decreasing computer costs, and improvements in computer literacy among health workers are creating enormous opportunities to use information technology to improve health systems and services by facilitating communication among the stakeholders of health care delivery and finance. Industries such as airlines, banking, and mobile technology have taken advantage of these opportunities to improve their efficiency and productivity, creating new levels of synergy and the ability for transactions to cross the boundaries of individual organizations. The authors believe that the time is ripe for a similar movement in health insurance and that a Health Data Dictionary (HDD) is a strong step that countries should take toward harmonizing information standards.

The authors developed this series to aid countries that are consolidating health insurance schemes to achieve universal coverage. Part 1 provides an overview for national policymakers on the role of the HDD and why establishing one early on is a key step in promoting system interoperability. The paper also offers practical advice on the steps that are required to make an HDD operational. Part 2 orients both policymakers and technical experts to some of the technical challenges and discusses approaches that can simplify HDD design, providing examples from countries that have previously developed HDDs and showing how they tackled similar issues. Part 3 introduces technical experts to an open source tool developed by PharmAccess Foundation that can be used by countries developing a national HDD. Countries can access this tool at www.jointlearningnetwork.org/content/tools. Depending on the interest and response from Joint Learning Network (JLN) members, this series may be extended to include additional topics in 2012.

The series was developed by Dr. Dennis J. Streveler, Senior HMIS Consultant to the World Bank and Professor, Medical Informatics, University of Hawaii, and Mr. Cees Hesp, Chief Technology Officer of PharmAccess Foundation. The topics were chosen based on observations from their respective field work for their organizations and direct needs expressed by the member countries of the Joint Learning Network. The JLN comprises 10 countries spread across Asia and Africa, all moving toward universal health coverage. Some JLN members had requested practical help in improving interoperability among disparate national health insurance schemes through the use of an HDD. The series was sponsored by the Rockefeller Foundation through funding for the Information Technology Track of the JLN. This series is a companion piece to the Collaborative Requirements for Health Insurance report prepared for the JLN and available under separate cover at www.jointlearningnetwork.org/content/tools.

The Role of the Health Data Dictionary

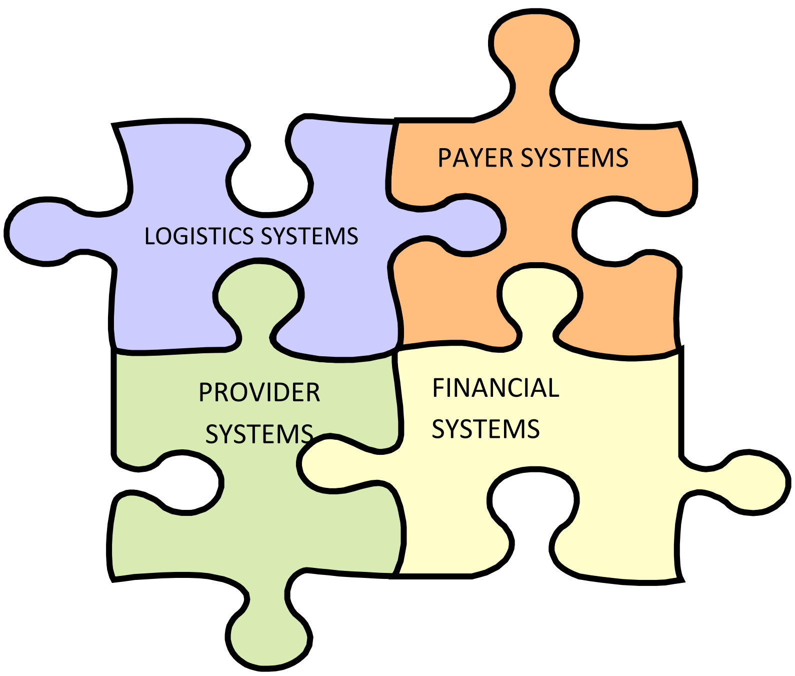
PART 1: HOW AN HDD PROMOTES INTEROPERABILITY AND A DISCUSSION OF THE POLITICAL PROCESS NEEDED TO NURTURE ONE

By Dennis Streveler, Professor, University of Hawaii

The need for "interoperability"

In the quest to attain universal health coverage (UHC), JLN member countries, like almost every other country in the world, struggle to "harmonize" or "integrate" their health information systems, including the discrete components that make up their health insurance information systems.

Figure 1. Components of health insurance information systems.



In the coming years, progress in health information systems will increasingly depend on the components listed in Figure 1 being able to "talk" with each other and share data. In technical parlance, we call this "interoperability"—recognizing that health insurance information systems are in fact a suite of applications that (must) work together. In practical terms, we mean that a Ministry of Health (or other organization) can derive increased synergy between applications, and thus an increased return on its investments (ROI) for systems that it has procured or built.

Just as humans require a common language by which to efficiently communicate, so do computers. While it is possible to soldier on, using translation techniques of various kinds (i.e., middleware, translation tables, mapping engines) to facilitate communication, this translation comes at a high price in terms of costs and delays and possible “mistranslations” of data. So too is it with computers, where middleware, translation tables, mapping engines, and all sorts of extra technological feats (at a high cost!) are needed to bridge communication between applications. And, as in the translation of natural language, the mistranslation of digital information also can be problematic and create serious repercussions downstream.

Systems must be able to “talk” the same “language” if they are to be able to cooperate in performing the processes of health insurance, and standards play an important role in defining vocabulary.

Some useful precedents in creating health data standards

The concept of creating health data standards and promoting a common language among practitioners is not new. There have been some notable global attempts to facilitate interoperability, the most important of which is arguably the creation of the International Classification of Diseases (ICD), which is now in version 10. This standard can trace its origins back to 1893.[1] Today the ICD is used almost everywhere in the world, and allows the relatively easy comparison of morbidity and mortality statistics between countries. If the ICD had not been invented and promulgated, we’d still be writing diagnoses in long hand! Now we take this landmark achievement for granted, but imagine a world where we allowed each caregiver to describe a patient’s condition in full text, in the native language of that country, in varying degrees of detail! [2]

Another effort was the emergence of the Health Level Seven (HL-7) standards, begun circa 1985, which tried to end the untenable situation of a patient needing to be registered separately in each information system within a hospital—upon admissions with the Hospital Information System (HIS), upon first lab test with the Laboratory Information System (LIS), upon first radiological exam with the Radiology Information System (RIS), etc. HL-7 was an important move forward as it allowed hospitals to “cherry pick” from available “HL-7 compliant” modules and still integrate them so they could work together in a reasonable way.

Despite the emergence of web-based systems, which inherently offer the opportunity for applications to more closely collaborate in their functioning, the existing standards are still woefully inadequate when it comes to processing insurance claims, helping with clinical decision-making, and doing the administrative work of health care planning and policymaking.

---

[1] Created and championed by the World Health Organization (WHO) as a way of assuring comparability of morbidity and mortality data from around the world.

[2] A good example is that clinicians might argue the difference between alternative ways of indicating an inflammation of the lungs—*pneumonia* and *pneumonitis*—which as character strings are considerably different. A single ICD code, however, nicely gets rid of the ambiguity.

Some countries, most notably Australia and Canada, have succeeded in creating significant standards within their own borders, but few major standards have emerged that have been embraced by other countries. This paper and the accompanying ones in this series provide an overview of the role of the HDD in promoting system interoperability and some tools and examples that may be useful as a country-level HDD effort is launched.

The role of the HDD

Linkages between providers and payers (and beneficiaries) are required for health insurance transactions. An HDD defines the transaction terms so that all parties can unambiguously understand exchanges between systems. Some examples of how an HDD can improve health insurance transactions include:

- **Promoting "clean" claims.** Can providers submit claims to payers that are readily understood by both sides so as to avoid many causes for the claims to be "rejected"?
- **Promoting "eclaims."** Can providers submit electronic versions of claims that computers on both sides can understand?
- **Streamlining "provider payments."** Can payers electronically route payments through the inter-banking system to providers and create remittance advice to justify the amount of the payment?
- **Resubmitting "rejected claims."** Can we make the process of resubmission of rejected claims simpler and easier?[3]
- Providing a simple method to answer the question "what is the status of my claim?[4] Can we make it easier to offer a claims status inquiry function that allows providers, and possibly beneficiaries, to know the status of each and every claim?

Outside health insurance, there are many other examples within the health sector of transactions between business stakeholders:

- **Connecting departments within a hospital.** The various modules of the hospital can communicate and work together without double-entry and without transposition as a user moves from module to module.
- **Connecting hospitals.** Hospitals may choose to use different hospital information systems (and electronic medical record systems) for a host of reasons—including size, organization, personal preference, and legacy experience. Can the systems still communicate when a patient is to be transported from one to another? In other words,

[3] Estimates of the cost of resubmitting a rejected claim vary from 2 to 10 times the cost of submitting the original claim!
[4] It is widely believed that a large percentage of "duplicate" claims happens when it is unclear what the status of the original claim is. It is just sometimes better therefore to submit yet another claim, just in case the first claim is unaccounted for! Of course this only adds to paper bloat and processing costs.

they still speak the same language even though they might have a very different "look and feel" and functionality.

- **Integrating logistics management.** Supply chains are long integrated processes that begin at manufacture/importation and end with the dispensing of a supply (or drug) to a patient in a hospital. Can we track that supply chain from beginning to end?

An HDD contains information about shared standards including a descriptive list of names (also called representations or displays), definitions, and attributes of data elements to be collected in an information system or database in the health sector. By standardizing definitions and ensuring consistency of use, the HDD enables conforming and comparable health information to be generated across the country, independent of the organization or system from which it originated.

The needed standards are compiled, published, and enforced in an HDD, which is simply a container to hold, organize, and disseminate the various standards and make them available to everyone developing and using e-health applications.

Creating the HDD

When a country decides to create an HDD, the first impulse is often to suggest adopting some other country's already existing HDD. Although looking at examples from other countries can inform the creation process, it can seldom replace it. Because each country has a unique perspective and its own legacy systems,[5] it is unlikely that the wholesale adoption of another country's HDD will be appropriate or successful.

Divide and conquer! – the segments of an HDD

The first task is to divide the HDD into segments—a sort of "divide and conquer" approach. The project team then identifies leaders for each of the various segments, each of which requires different subject matter expertise. The leadership and cooperation of these experts are enormously important for ensuring the credibility and ownership of the resultant HDD.

How many segments to include in an HDD is a matter of discussion. For the first iteration of an HDD, it is generally advised to keep the number of segments relatively small, perhaps ten. Later, as more systems are built, that number will naturally increase as new domains arise in the process. Table 1 provides a possible first set of segments to consider.

---

[5] "Legacy systems" are those already in place in a country. It is usually impractical, and unwise in any case, to replace all the existing systems with standards-compliant ones. Thus, the existing systems will need to be "retrofitted" to accommodate the new standards. What new standards you choose will determine how difficult that task will be.

Table 1. Possible HDD segments

SEGMENT	CONSIDERATION(S)
1. Patient demographics	How will each person be identified and grouped. What is the “Unique Patient Identifier”?
2. Patient history	The aspects of a patient’s history that will be connected. In the simplest form, this might be limited to the list of chronic disease categories to which this patient belongs. (Note: This later becomes the basis for the electronic patient record.)
3. Providers	Each licensed, credentialed caregiver needs to have a unique Provider ID number.
4. Facilities	Each licensed, credentialed health venue needs to have a unique Facility ID number.
5. Health insurance	This segment will depend to a large extent on the sophistication of the health insurance scheme—whether it has different benefit plans, options for dependents status, different rates for provider reimbursement, and so on.
6. Health resources	Including human resources and consumables (pharmaceuticals and supplies).
7. Utilization management	How will we know if health resources are being used wisely and productively?
8. Quality management	What quality measures will you adopt to measure the outcomes of the health care delivery system?
9. Financial management	How will budgeting and accounting be “standardized” so that we know the financial status?
10. Clinical protocols and guidelines	Where will clinical standards be amassed and disseminated?

An expert task force, or work group, representing each segment of the HDD will need to be formed to suggest appropriate standards, state them clearly, and seek validation and concurrence from agencies concerned with that segment. It is crucial to recruit the right people for this duty and to give them sufficient time and resources to concentrate on the task and seek input from peers. The people who know the processes best can make the biggest contribution toward the HDD. However, these people are typically involved in many other activities, and it is possible that the difficulty of this task will be underestimated

The process of creating an HDD requires considerable energy, stamina, and compromise. Many arguments will arise about which potential standard is "better" than another. Although there is some reason to say that one choice might be better than another (e.g., choosing to use ICD-10 is probably a better and more strategic decision than using ICD-9 at this time***), the key goal is for everyone to agree to one standard.***[6] ***It can be said that almost any standard is better than no standard!***

This work will need to be facilitated and encouraged by the highest levels of the Ministry of Health (MOH). Without executive sponsorship, the working groups can easily descend into endless discussions bearing little fruit. The leaders of the MOH need to issue a firm directive that an HDD will be created and indicate when the first draft is to be completed. The rate at which countries complete this task varies widely, but ***at least 6 months*** appears to be the minimum to complete a first draft. (In one sense, an HDD is never completed because it will continue to be expanded and corrected for decades to come.)

Although the specific steps taken to create an HDD will vary from country to country, most countries typically follow a six-step process:

1. Review existing de facto standards used by some legacy systems.
2. Review some other countries' standards, especially those most similar to your own.
3. Gain consensus within the MOH on the importance of this activity.
4. Launch the "Create our HDD" activity, appointing a strong overall coordinator to lead the day-to-day tasks.
5. Prioritize the work of the activity, choosing to address those segments that are most needed to move ahead with plans for new e-health applications.
6. Create working groups as appropriate to report recommendations for each segment. Workgroup members might be found in the MOH, public health institutes, health insurance organizations, university medical schools, medical specialty boards, hospital syndicates, or other organizations. While it is important to be inclusive in this process, work groups should remain manageable, with members being accountable to the overall outcome.

[6] There are some who say *any* standard is better than *no* standard, and there is some truth to this approach.

Note: Parat 2 of this series suggests methods to record the output of the working groups, and Part 3 provides a tool for countries to use when creating a form that can be read by both humans and computers (XML).

Publishing and promoting the HDD

As the HDD is being created, the working group should keep in mind that the HDD needs to be accessible and easily referenced by all stakeholders on an ongoing basis, including:

- Software developers, applications vendors, and software project implementers.
- Institutions contemplating purchasing e-health applications "off-the-shelf."
- International organizations that might sponsor the purchase or development of these applications.
- Statistical bureaus that interpret health data.

Although distributing printed copies of the HDD might be appropriate in some settings, web-based publishing of the HDD is more effective. This allows the HDD to be easily searched, rearranged, and sorted, and extracts can be copied into other documents.[7] Most importantly, electronic distribution ensures that the reader can always access the ***latest*** version of this dynamic document.

Once published, the HDD has to be actively promoted—otherwise there is the danger that the HDD could simply languish "on the shelf" rather than being used to streamline communications between processes.

Enforcing the HDD

The most effective HDDs are sponsored directly by a high authority. An HDD that is tepidly enforced is not an HDD at all! The effort can be considered successful only if the HDD is used to improve interoperability.

To accommodate real-world concerns about the impact of the new HDD on information systems strategy, leaders must allow sufficient time for the new standards to take effect. Typically, a high-level authority will issue a statement such as:

> Beginning on January 1, 201x, in the Republic/Kingdom of xx, all new health information systems components will be built following the health information standards contained in the National Health Data Dictionary (dated xx/xx/xx, version x). Similarly, all new health information systems to

[7] An example of this, from Australia, can be found at http://meteor.aihw.gov.au/content/index.phtml/itemId/268110.

be procured after January 1, 201x, will be compliant with said dictionary. Finally, beginning on January 1, 201x, all legacy systems will have been retrofitted, converted, or replaced by compliant modules. Any exceptions to this order must be approved in writing by the office of the Minister of Health.

It will not be easy to gain acceptance of the new standards because there are many vested interests in the existing "standards" or, should we say, lack of standards. For example:

- Proprietary vendors sometimes prefer to use their own "proprietary standards" to gain marketplace advantage.

- Some stakeholders may complain that the time period for compliance is too short. They will likely delay until the applicable deadline looms and will continually ask for extensions.

- Most problematically, there will be a question about what to do with systems in development at the time the new standards are published. Despite the inconvenience and cost of forcing the new application to conform to the new standards, this is often preferable to retrofitting them later, when changes would also result in substantial costs.

- Finally, private-sector health institutions may not wish to participate. This can be a serious political issue to address. Losing control of data from the private sector (especially if it represents a sizeable piece of the overall health sector) means never being able to see "the whole picture" of health care in a country. Usually, though, government health officials have some leverage in this struggle, such as the ability to mandate that private providers follow the HDD (at least the important segments of it) in order to be eligible to receive funds from the public health insurance schemes.

Enforcement is **crucial** to making the transition and making it possible for the country to progress to the next level of system development, which involves "integration" and "interoperability" of the various e-health applications. Without agreed-upon standards that are documented in one location, the country will not have a firm foundation upon which to build its e-health applications.

Going a step further, some countries have initiated a process to certify compliance of information systems to the HDD standard. When this process is used, only certified systems can be sold to health institutions. The process ensures that certified modules can be "plugged in" to your e-health application matrix with a minimum of effort.

Updating the HDD

Updating the HDD is a continual and never-ending activity. Policy changes precipitate new data items or applications. New disruptive technologies will suggest better ways to perform a particular task. New medical procedures will be added, health reforms will be rolled out, and so on.

A host organization should be identified that can manage this ongoing activity and institute a process through which requests for updates are collected and considered for their validity and advisability, can make the appropriate changes, and can notify stakeholders that new versions are available (and when they will become effective). This work requires good version control of the HDD, not allowing it to change continuously or capriciously. Implementing changes too frequently can lead to chaos if the health system cannot react quickly enough to keep itself aligned to the latest changes. Implementing changes too infrequently will probably lead to an HDD that is "out of date" and therefore easily discredited. At the very least, a process of HDD "versioning" is recommended. Although versions might initially be released on a more-frequent basis (for example, quarterly), less-frequent releases (for example, annually) might be appropriate in later stages once the HDD "settles down" and gets "shaken out."

Institutions that might be good choices for updating the HDD include the statistical bureau of the MOH, a public health institute, or a leading medical school or teaching hospital. It is wise to choose an organization that not only has the capacity but also exhibits leadership in the health sector. The choice largely depends on the organogram of the country's health sector. The selected institution must be able to update the HDD in a thoughtful and timely manner and must continue to focus attention over the long term on the importance of having a current HDD.

Conclusion

Managing the HDD, from concept to reality, is a challenging but necessary activity if a country wishes to realize the dream of an ***integrated*** and ***harmonized*** health information system. Although some of the tasks surrounding the HDD may sound daunting, either technically or politically, or both, this essential work need not be so burdensome. Start with those pieces of the HDD that need to be done *now* and add segments to the HDD in the future as your information system needs grow and new systems are being developed or procured. Developing the HDD will enable your health information systems components to more easily "talk to each other" and help you leverage the information in the health sector more effectively.

Primer on Creating a Health Data Dictionary

PART 2: EXAMPLES OF HDDS IN USE AROUND THE WORLD

By Cees Hesp, Chief Technology Officer, PharmAccess Foundation

Overview

Part 1 of this series describes the importance of having a health data dictionary (HDD) as the foundation for interoperability, a standardized language to facilitate the exchange of data and transactions across health insurance information systems.

On a technical level, creating an HDD can be quite challenging, as illustrated by the excerpt in Figure 1 from a white paper by 3M, a commercial vendor of data dictionary software.

Figure 1. Different definitions of the word "cold."

What does the word "cold" mean?

- An accident victim brought into the emergency room tells the attending physician, "I feel cold."
- A pulmonologist tells a 58-year-old male patient that he is suffering from COLD—chronic obstructive lung disease.
- You call your family practitioner for an appointment and tell the receptionist, "I have a bad cold that's not getting better."

As countries grapple with building an HDD, Part 2 of this series provides examples of HDDs in use around the world. It provides a starting point for people tasked with creating an HDD, focusing on the technical, practical, and content details, and drawing examples from work done in a number of countries/territories, including Australia, Ghana, the Netherlands, New Zealand, Palestine, Tanzania, and the United States. This part is meant as a "primer" in the sense that both good and "bad" examples are shown, and it is left up to readers to make up their own minds and follow the example that best suits their needs.

Management summary

The process of creating an HDD will vary from country to country, but there are best practices that can serve as a guide. The American Health Information Management Association published an article in 2006, "Guidelines for developing a data dictionary," that provides a useful starting point (Figure 2), although it is geared toward the American context of developing systems for electronic health records. The full article can be downloaded from the association's website: www.ahima.org.

Figure 2. Excerpt from "Guidelines for developing a data dictionary."

1. Design a plan for the development, implementation, and continuing maintenance of the data dictionary.
2. Develop a data dictionary that integrates common data elements used across enterprises.
3. Ensure collaborative involvement and buy-in of all key stakeholders when data requirements are being defined for an information system.
4. Develop an approvals process and documentation trail for all initial data dictionary decisions and for ongoing updates and maintenance.
5. Identify and retain details of data versions across all applications and databases.
6. Design flexibility and growth capabilities into the data dictionary so that it will accommodate architecture changes resulting from clinical or technical advances or regulatory changes.
7. Design room for expansion of field values over time.
8. Follow established guidelines or rules for metadata registry (data dictionary) construction to promote interoperability and automated data sharing.
9. Adopt nationally and internationally recognized standards and normalize field definitions across datasets to accommodate multiple end-user needs.
10. Beware of differing standards for the same clinical or business concepts.
11. Use geographic codes and geocoding standards.
12. Test the information system to demonstrate conformance to standards as defined in the data dictionary.
13. Provide ongoing education and training for all staff as appropriate to their use of data elements and their definitions.
14. Assess the extent to which the use of the agreed-upon data elements supplies consistency of information sharing and avoids duplication.

Source: American Health Information Management Association (AHIMA). Guidelines for developing a data dictionary. *Journal of AHIMA*. 2006; 77(2):64A–D.

International standards

There are international standards on how to describe data elements, although it should be noted that these standards are not widely used outside Australia and the United States. These standards are:

- ISO/IEC Standard 11179 Specification and Standardization of Data Elements.
- ISO/IEC International Standard 11179-3:2002 (Information Technology-Metadata Registries-Part 3: Registry metamodel and basic attributes).

The Australian national data dictionaries use the following language to describe its format and structure, based on the ISO/IEC Standard 11179 (Figure 3).

Figure 3. Excerpt on format and structure of the Australian national data dictionaries.

> All definitions in the national data dictionaries are presented in a standard format based on ISO/IEC International Standard 11179-3:2002 (Information Technology-Metadata Registries-Part 3: Registry metamodel and basic attributes). This is the international standard for defining data elements issued by the International Organization for Standardization and the International Electrotechnical Commission. Collectively, the format describes a set of attributes for data definitions.
>
> The application of this international standard across data dictionaries in the health, housing, and community services sectors adds to the completeness, integrity, and consistency of data definitions and consequently to the quality and utility of national data.

Objectives and information needs

The content of an HDD should be driven by what needs to be communicated within a health system. For example, the creators of the Australian HDD stated its objectives as follows (Figure 4):

Figure 4. Australian objectives of having a national HDD.

> The Dictionary is designed to improve the comparability of data across the health field. It is also designed to make data collection activities more efficient by reducing duplication of effort in the field, and more effective by ensuring that information to be collected is appropriate to its purpose.
>
> The objectives of the *National Health Data Dictionary* are to:
>
> - establish a core set of uniform definitions relating to the full range of health services and a range of population parameters (including health status and determinants);
> - promote uniformity, availability, reliability, validity, consistency and completeness in the data;
> - accord with nationally and internationally agreed protocols and standards, wherever possible;
> - promote the national standard definitions by being readily available to all individuals and organisations involved in the generation, use and/or development of health and health services information;
> - facilitate and promote the development of good data definitions across the health sector.
>
> The National Health Data Committee is responsible for coordinating the development and revision of the *National Health Data Dictionary*.

HDDs are living documents

As Part 1 emphasized, HDDs are living documents that evolve over time. As an example: Australia's first HDD was published in 1989 as the "National Minimum Data Set—Institutional Health Care" and is now at version 15. This means a new version has been published every 18 months, on average. The differences between versions are quite substantial (Figures 5a and 5b).

Figure 5a. Summary of changes in Australia's national HDD, version 12 (2003).

Version 12 comprises:

- the addition of 87 new data elements
- the modification of 25 data elements requiring a version change
- the modification of 13 data elements that did not require a version change
- the retirement of 3 obsolete data elements concepts
- the introduction of a new NMDS for Non-admitted patient emergency department care
- changes to the NMDS for Alcohol and other drug treatment services
- three new types of metadata set; Data Set Specifications (DSS) for Cardiovascular disease (clinical), Diabetes (clinical) and Health care client identification.

Figure 5b. Summary of changes in Australia's national HDD, version 15 (2010).

Table 1: Summary table of updates to the NHDD since version 14

Registration status	National minimum data sets	Data set specifications	Data clusters	Data elements	Classifications	Glossary items
Standards (new)	Nil	13	33	122	1	1
Standards (revised)	20	8	6	84	7	1
Superseded	20	8	6	86	7	1
Retired	Nil	Nil	Nil	Nil	Nil	Nil

Whereas a bulleted list was used in 2003 to enumerated the changes (Figure 5a), this section was replaced by a tabular format in 2010 (Figure 5b).

Creating an HDD

Creating an HDD is much like creating a database and starts with asking basic questions about the information needed to facilitate communication within a health insurance information system.

- Which objects need to be modeled (patients, providers, payers, etc.)?

- Which properties or attributes of these objects need to be recorded (name, date of birth, gender, etc.)?
- What are the existing relationships among these objects? For example, does a patient visit different providers over time?
- Which "business rules" govern the allowed behavior of these objects and properties?
- How much history needs to be retained? For example, should the system store information about a patient's current address as well as previous addresses?

The objects will usually translate into tables in a relational database, and the properties or attributes usually translate into columns. The specific tables or objects required will depend on the particular goals and priorities of your HDD.

The Palestinian Health Data Dictionary, 2nd Edition (2005), offers one example of how to organize the data into segments. The country's HDD consists of the following segments:

- Health care provider
- Person information
- Health event
- Communicable diseases
- Chronic diseases

The Palestinian data model, therefore, likely includes tables/objects focused on "provider," "person," "health event," and "diseases." (Apparently at that time they had no immediate need for "insurer/payer.")

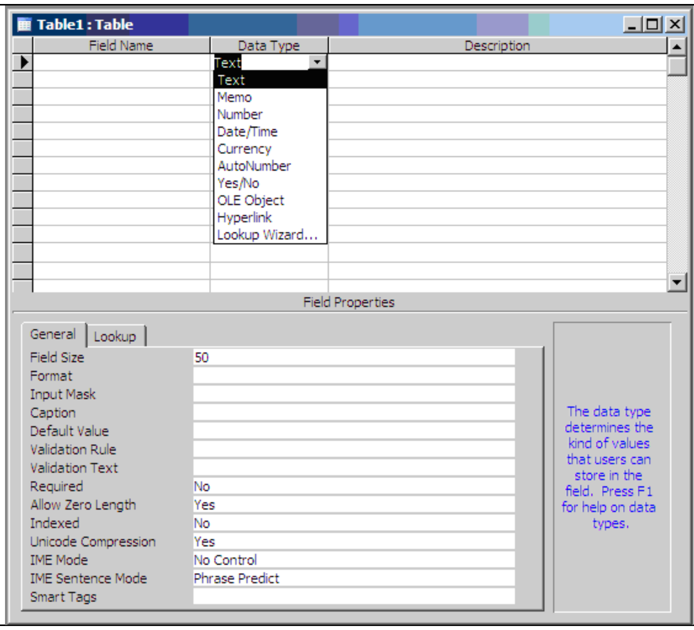
A practical method of identifying which objects and properties to include is to examine the paper forms currently exchanged between providers and payers in your country. This approach is further described in the section on e-claims.

Which data to collect/store per data element

Taking a relatively simple database system such as Microsoft Access as an example (Figure 6), a new table requires the following details for each column:

Figure 6. Details for creating a table in Microsoft Access.

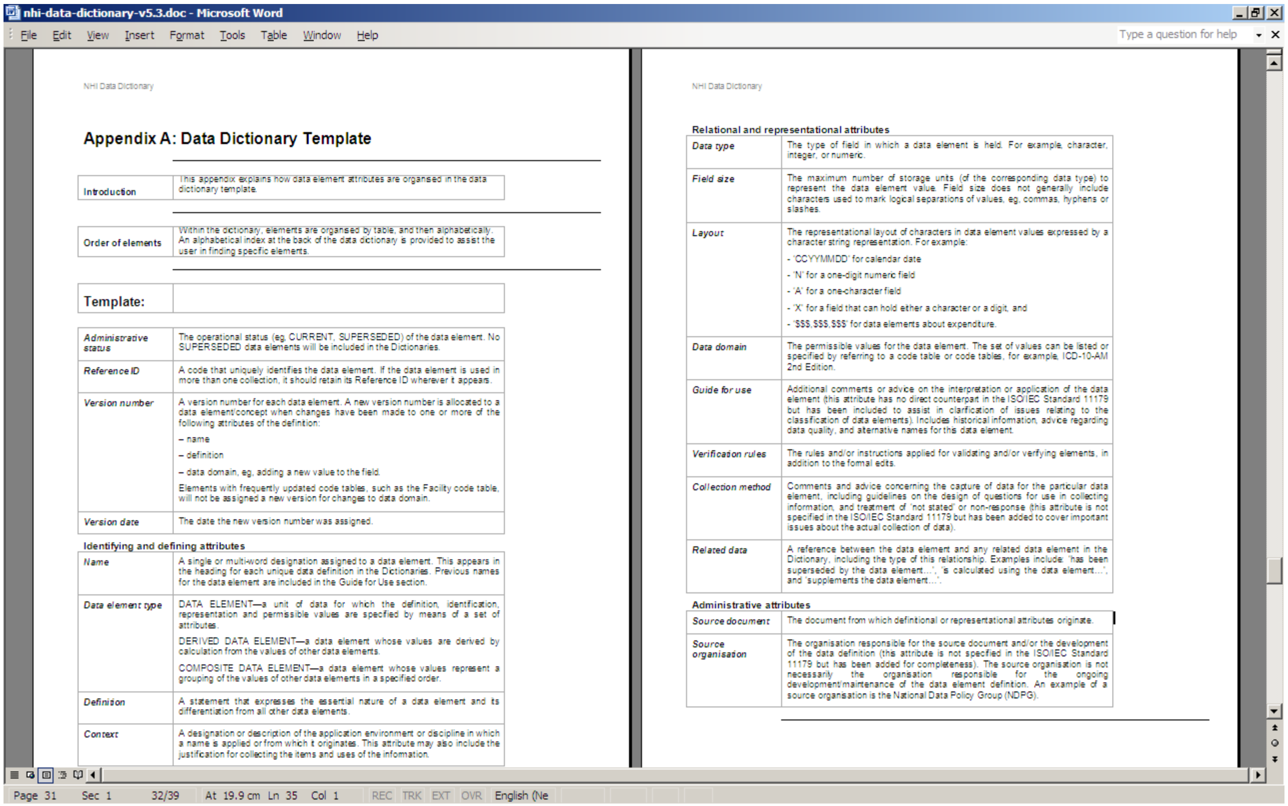
- Field name.
- Data type (number, text, date).
- Field size (i.e., length).
- Default value.
- Validation rule.
- Required yes/no (i.e., can it be left empty?).
- Indexed (i.e., does the value have to be unique, as in the case of unique identifier?)



This is the same kind of information commonly found in HDDs, although the columns/properties are usually called "data elements" or "data items."

As an appendix to its National Health Index Data Dictionary, New Zealand provides a useful template for the types of information to include in the description of data elements (Figure 7).

Figure 7. Template from New Zealand describing what to include in a data dictionary.



NHI Data Dictionary

Appendix A: Data Dictionary Template

Introduction	This appendix explains how data element attributes are organised in the data dictionary template.
Order of elements	Within the dictionary, elements are organsed by table, and then alphabetically. An alphabetical index at the back of the data dictionary is provided to assist the user in finding specific elements.

Template:	
Administrative status	The operational status (eg, CURRENT, SUPERSEDED) of the data element. No SUPERSEDED data elements will be included in the Dictionaries.
Reference ID	A code that uniquely identifies the data element. If the data element is used in more than one collection, it should retain its Reference ID wherever it appears.
Version number	A version number for each data element. A new version number is allocated to a data element/concept when changes have been made to one or more of the following attributes of the definition: – name – definition – data domain, eg, adding a new value to the field. Elements with frequently updated code tables, such as the Facility code table, will not be assigned a new version for changes to data domain.
Version date	The date the new version number was assigned.

Identifying and defining attributes

Name	A single or multi-word designation assigned to a data element. This appears in the heading for each unique data definition in the Dictionaries. Previous names for the data element are included in the Guide for Use section.
Data element type	DATA ELEMENT—a unit of data for which the definition, identification, representation and permissible values are specified by means of a set of attributes. DERIVED DATA ELEMENT—a data element whose values are derived by calculation from the values of other data elements. COMPOSITE DATA ELEMENT—a data element whose values represent a grouping of the values of other data elements in a specified order.
Definition	A statement that expresses the essential nature of a data element and its differentiation from all other data elements.
Context	A designation or description of the application environment or discipline in which a name is applied or from which it originates. This attribute may also include the justification for collecting the items and uses of the information.

NHI Data Dictionary

Relational and representational attributes

Data type	The type of field in which a data element is held. For example, character, integer, or numeric.
Field size	The maximum number of storage units (of the corresponding data type) to represent the data element value. Field size does not generally include characters used to mark logical separations of values, eg, commas, hyphens or slashes.
Layout	The representational layout of characters in data element values expressed by a character string representation. For example: - 'CCYYMMDD' for calendar date - 'N' for a one-digit numeric field - 'A' for a one-character field - 'X' for a field that can hold either a character or a digit, and - '$$$,$$$,$$$' for data elements about expenditure.
Data domain	The permissible values for the data element. The set of values can be listed or specified by referring to a code table or code tables, for example, ICD-10-AM 2nd Edition.
Guide for use	Additional comments or advice on the interpretation or application of the data element (this attribute has no direct counterpart in the ISO/IEC Standard 11179 but has been included to assist in clarification of issues relating to the classification of data elements). Includes historical information, advice regarding data quality, and alternative names for this data element.
Verification rules	The rules and/or instructions applied for validating and/or verifying elements, in addition to the formal edits.
Collection method	Comments and advice concerning the capture of data for the particular data element, including guidelines on the design of questions for use in collecting information, and treatment of 'not stated' or non-response (this attribute is not specified in the ISO/IEC Standard 11179 but has been added to cover important issues about the actual collection of data).
Related data	A reference between the data element and any related data element in the Dictionary, including the type of this relationship. Examples include: 'has been superseded by the data element…', 'is calculated using the data element…', and 'supplements the data element…'.

Administrative attributes

Source document	The document from which definitional or representational attributes originate.
Source organisation	The organisation responsible for the source document and/or the development of the data definition (this attribute is not specified in the ISO/IEC Standard 11179 but has been added for completeness). The source organisation is not necessarily the organisation responsible for the ongoing development/maintenance of the data element definition. An example of a source organisation is the National Data Policy Group (NDPG).

Country examples

What a health data dictionary looks like and what information it commonly contains is best explained by looking at some real-life examples.

Australia

To see how local needs and approaches can impact the evolution of an HDD, consider the experience of Queensland, Australia. Figure 8 shows a data element called "surname" in the state's HDD published in 1998. By 2003, a revised version contained the same data element, now called "family name," with much more detailed information (Figure 9). The elements have the same ID, 040002, but a different version number.

Figure 8. Entry for “surname” from 1998 Queensland health data dictionary, p. 273.

Surname			
Data Element ID	040002	Version number	1
Type	DATA ELEMENT		
Status	CURRENT	(Date)	15/01/1998
Definition	Surname of person		
Context	Institutional health care		
Data Type	Character		
Representational Form	TEXT	Representation Layout	A(50)
Minimum Size	1	Maximum Size	50
Data Domain	Left justified character string		
Guide for Use			
Verification Rules			
Related Data References	-is used in conjunction with First name, QHLTH 040003 version 1 -is used in conjunction with Second name, QHLTH 040004 version 1 -relates to the data element Family name, QHLTH 040002 version 2		
Source Document			
Source Organization			
Comment			

Figure 9. Entry for “family name” from 2003 Queensland health data dictionary, pp. 135–136.

Family Name			
Data Element ID	040002	Version number	2
Type	DATA ELEMENT		
Status	CURRENT	(Date)	02/09/2003
Definition	That part of a name a person usually has in common with some other members of his/her family, as distinguished from his/her given names.		
Context	Administrative purposes and individual identification.		
Data Type	Alphanumeric		
Representational Form	TEXT	Representation Layout	AN(40)
Minimum Size	1	Maximum Size	40
Data Domain	Text		
Guide for Use	The agency or establishment should record the client’s full “family name” on their information systems. NCSDD specific: In instances where there is uncertainty about which name to record for a person living in a remote Aboriginal or Torres Strait Islander community, Centrelink follows the practice of recording the indigenous person’s name as it is first provided to Centrelink. In situations where proof of identity is required, the name recorded should appear on a majority of the higher point scoring documents that are produced as proof of identity.		
Verification Rules			
Related Data References	-Relates to the data element concept Person name, QHLTH 040857 version 1 -Relates to the data element Given name, QHLTH 040003 version 1 -Relates to the data element Name conditional use flag, QHLTH 040853 version 1 -Relates to the data element Name suffix, QHLTH 040851 version 1		

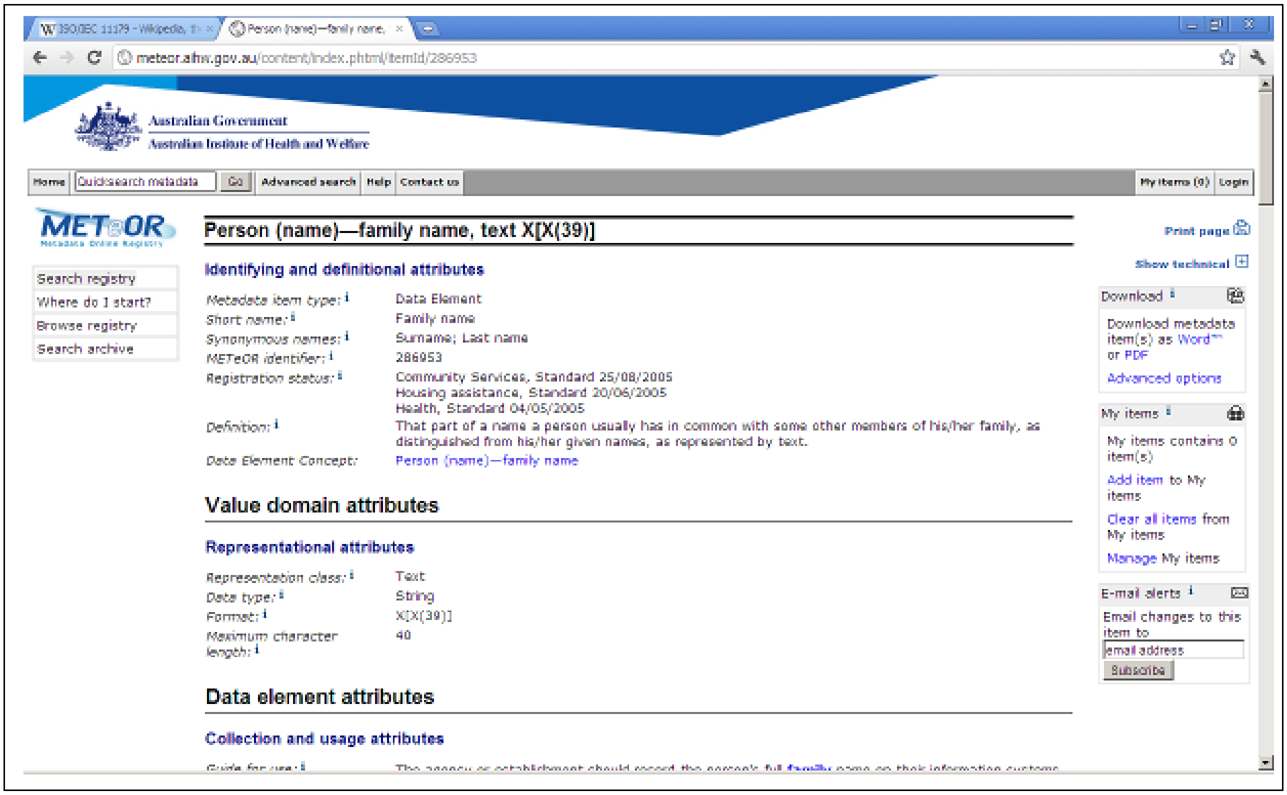
	-Relates to the data element Name title, QHLTH 040850 version 1 -Relates to the data element Person name type, QHLTH 040858 version 1 -Relates to the data element Surname, QHLTH 040002 version 1
Source Document	National Health Data Dictionary
Source Organization	Australian Institute of Health and Welfare
Comment	This metadata item is common to both the National Community Services Data Dictionary and the National Health Data Dictionary. Often people use a variety of names, including legal names, married/maiden names, nicknames, assumed names, traditional names, etc. Even small differences in recording—such as the difference between MacIntosh and McIntosh—can make record linkage impossible. To minimise discrepancies in the recording and reporting of name information, agencies or establishments should ask the person for their full (formal) “Given name’ and “Family name’. These may be different from the name that the person may prefer the agency or establishment workers to use in personal dealings. Agencies or establishments may choose to separately record the preferred names that the person wishes to be used by agency or establishment workers. In some cultures it is traditional to state the family name first. To overcome discrepancies in recording/reporting that may arise as a result of this practice, agencies or establishments should always ask the person to specify their first given name and their family name or surname separately. These should then be recorded as “Given name’ and “Family name’ as appropriate, regardless of the order in which they may be traditionally given. NCSDD specific: Selected letters of the family name in combination with selected letters of the “Given name’, “Date of birth’ and “Sex’ may be used for record linkage for statistical purposes only. Other References: AS4846 Health Care Provider Identification, 2004, Sydney: Standards Australia

Most of the informational items that were left empty in the first version were later filled out in great detail. While the initial focus was on the formal or technical aspects, it became clear later that the context in which these definitions are used is also important. Having definitions without someone actively managing, maintaining, and owning them is asking for trouble.

Once an HDD has been established, making even small changes involves both effort and cost. For example, changing the maximum length of some data element from 50 to 40 characters requires significant work by programmers, database administrators, and the designers of paper and electronic forms. Changing “surname” to “family name” is even more significant. Imagine that your country has a postal code and that overnight it has to be called “zip code” in all official correspondence.

As more and more definitions were added, the Australian HDD became so comprehensive that it was split into two volumes. When even that became unmanageable, it was replaced by a dedicated website, METeOR, short for Metadata Online Registry (Figure 10).

Figure 10. Metadata Online Registry website (METeOR).



From this website, the paper version of the HDD (ISBN 978-1-74249-050-2) can be downloaded in PDF format at more than 3,000 pages. One can also query the HDD interactively or programmatically (more about this later).

New Zealand

New Zealand clearly borrowed from its neighbor Australia but also added aspects of its own, (e.g., listing "last name" and "surname" as synonyms for "family name"). Note the formal indication that this data element is considered mandatory (Figure 11).

Figure 11. Family name as defined in New Zealand's National Health Index (version 5.3).

Family name

Administrative status

Reference ID:	A0013	*Version:*	1.0	*Version date:*	14-Oct-2003

Identifying and defining attributes

Name:	Family name
Name in database:	family_name
Other names:	Last name, Surname
Element type:	Data element
Definition:	The family name (surname) of a healthcare user.
Context:	

Relational and representational attributes *Mandatory*

Data type:	char	*Field size:*	25	*Layout:*	
Data domain:					
Guide for use:	With First name, Sex, and Date of birth, this provides a key for searching for the record.				
Verification rules:					
Collection method:	May only include upper- and lower-case alphabetic characters, spaces, hyphens and apostrophes.				
Related data:	First given name Second given name Third given name				

Administrative attributes

Source document:	
Source organisation:	

Palestine

The Palestinian Health Data Dictionary, 2nd edition, April 2005, takes a "minimalist" approach similar to that of New Zealand (Figure 12). The technical part is at the top, and the "usage" part is at the bottom.

Figure 12. Family name not formally defined in Palestinian Health Data Dictionary (2nd edition, April 2005).

Patient Name	
Data Element Name	Patient Name
Reference ID	B003
Data Element Type	Data Element
Definition	The name of patient
Data Type	Alpha
Field Size	80

Usage and Content: This data element is essential because of its value in probabilistic linkage, both as a linking variable as well as a confirmatory variable to determine appropriate linkage. It is recognized that this data element requires careful protection from misuse, but it is more appropriate to regulate appropriate use of this field.

Codes and Values: None

Technical Comments: It should be coded in a multiple fields, then the format of this fields should be FIRST (first name), SECOND (Father's name), THIRD (Grandfather's name), LAST (Family name).
The FIRST, SECOND and LAST must be coded as an Essential, the THIRD as a Desirable.

Notable differences:

- The Australians changed "Surname" to "Family Name" and shortened the length from 50 characters (original length) to 40 characters. Initially, the Australian specification did not allow for numbers in names, cf. A(50) versus AN(40), and did not formally indicate if an element was mandatory (only loosely described in "Guide for Use" as "the agency should record...").
- The national heath index used in New Zealand specifies 25 characters.
- The Palestinians do not have a formal definition of "Family Name" but use (a maximum of) 80 alphanumeric characters to record "Patient Name."

Examples of code tables

In an HDD, many data elements will have or belong to a "domain" consisting of predefined, coded values. The simplest example of such a domain is sex (male/female). Even here, however, different representations have been implemented:

- M = Male, F = Female, U = Unknown, X = Unspecified
- 1 = Male, 2 = Female, 9 = Unknown, 99 = Unspecified

Below are three examples of different coding schemes used. Note that:

- Example 1 uses a simple numbering system consisting of integers, where the numbers themselves do not have any meaning but only serve to be unique.
- Example 2 uses so-called “mnemonics” (i.e., codes that are easy to remember).
- Example 3 has gaps in the numbering, suggesting that there is some internal logic in the different categories being described and/or that room has been left for future additions. Note also that this example has some rather cryptic descriptions and that some of the meanings seem to partly overlap.

Example 1: Accommodation type

Definition: The type of accommodation setting in which a person usually lives/lived, as represented by a code.

Value	Meaning
1	Private residence (e.g., house, flat, bedsitter, caravan, boat, independent unit in retirement village), including privately and publicly rented homes
2	Psychiatric hospital
3	Residential aged care service
4	Specialized alcohol/other drug treatment residence
5	Specialized mental health community-based residential support service
6	Domestic-scale supported living facility (e.g., group home for people with disability)
7	Boarding/rooming house/hostel or hostel type accommodation, not including aged persons’ hostel
8	Homeless persons’ shelter
9	Shelter/refuge (not including homeless persons’ shelter)
10	Other supported accommodation

Source: Australian Health Data Dictionary, version 15, 2010, p. 1695.

Example 2: Health care provider type

Definition: Types and levels of health services provided by health sectors.

Value	Meaning
P0	Primary health care
P1	Level I
P2	Level II
P3	Level III
P4	Level IV
H0	Hospital
H1	Peripheral
H2	Central
H3	Specialist

Source: Palestinian Health Data Dictionary, 2nd edition, 2005, p. 22.

Example 3: Dosage form
Definition: None given (the form in which drugs are prescribed).

Value	Meaning	Value	Meaning
01	ENEMA	47	DIALYSIS SOLN.
02	DEVICE	51	OINT.
11	VIAL	52	CREAM
12	AMP.	53	JELLY
13	CART	54	E/OINT
21	TAB.	55	LOTION
22	CAP.	56	GEL/LOTI.
23	S.G.C.	57	CR/OINT
31	SYR.	61	SUPP.
32	SUSP.	62	VAG.TAB.
33	ELIXIR	63	OVULE
41	LIQUID	71	DROPS
42	SOLN.	72	E/DROPS
43	I.V SOLN.	81	AERESOL
44	EMULSION	82	INHAL
45	TOPICAL SOLN.	91	POWDER
46	SCRUB		

Source: Palestinian Health Data Dictionary, 2nd edition, 2005, p. 304.

Alternative to "paper" dictionaries: XML

While having a paper dictionary is useful to start the exercise of standardization, the dictionary will quickly reach a size that will make it difficult to use and maintain.

XML, short for Extensible Markup Language, has become the de facto standard for exchanging information between disparate information systems. It can be used to send messages from one computer system to another and to describe any metadata structure. For example, it can be used to describe what a data element in an HDD looks like.

XML basics
XML is a so-called "tag-based" language. A tag is a (preferably) meaningful term, written between angle brackets (< >). Each opening tag must be matched by a closing tag. Figure 13a shows a very simple "person" record. The opening tag <person> on line 1 is closed by </person> on line 5. Note the forward slash (/) in the closing tag.

Figure 13a. Simple "person" record, with dots indicating where the data go.

```
01 <person>
02    <firstName>...</firstName>
03    <lastName>...</lastName>
04    <dateOfBirth>...</dateOfBirth>
05 </person>
```

This template can be filled out by substituting real data for the dots, producing Figure 13b below. This describes that there is a person called John Doe, born on December 12, 1980.

Figure 13b. Simple “person” record filled out for John Doe.

```
<person>
   <firstName>John</firstName>
   <lastName>Doe</lastName>
   <dateOfBirth>1980-12-12</dateOfBirth>
</person>
```

To describe that there is a family consisting of two people, we could have the format shown in Figure 13c:

Figure 13c. “Person” records combined to form a “family.”

```
<family>
   <person>
      <firstName>John</firstName>
      <lastName>Doe</lastName>
      <dateOfBirth>1980-12-12</dateOfBirth>
   </person>
   <person>
      <firstName>Mary</firstName>
      <lastName>Doe</lastName>
      <dateOfBirth>1981-11-11</dateOfBirth>
   </person>
</family>
```

Several things are worth noting:

- XML is called “extensible” because it does not prescribe which tags should be used. Rather, new tags can be added as needed as long as all communication partners have a common understanding of what the tags mean.
- Tags can be “nested” to represent different hierarchical relationships. In Figure 13c, for example, <person> is nested within <family>, and <dateOfBirth> is nested within <person>. The relationship of <dateOfBirth> to <person> is something like “property of,” whereas the relationship between <person> and <family> is something like “member of.”
- Tags cannot have spaces in their names. This is why one often sees “camel case” being used (i.e., each new word starts with a capital letter except the first word). In XML, <firstName> could also be written as <first_name> or <FirstName>, whichever “standard” the communication partners consistently follow. Because of XML’s case-sensitivity, the exact name and spelling of tags is one of the first things that communication partners must agree upon.

XML documents tend to be rather verbose because of the repeating pairs of opening and closing tags. Fortunately, a shorthand notation is also available (Figure 13d).

Figure 13d. Shorthand notation using XML attributes rather than XML elements.

```
<family>
   <person firstName="John" lastName="Doe" dateOfBirth="1980-12-12" />
   <person firstName="Mary" lastName="Doe" dateOfBirth="1981-11-11" />
</family>
```

Using XML for metadata

XML can also be used to describe metadata—data about data. This is essentially what a data dictionary does. It provides data (e.g., definition, description, guidelines for use) about the data items in the dictionary.

Figure 14 shows a potential XML representation of the information described previously in the example from the Queensland data dictionary. Note the use of the word “potential”—other representations could be used as well, and would be equally valid.

Figure 14. XML representation of the information contained in Figures 8 and 9.

```
<dataElement id="..." version="...">
   <dataElementName>...</dataElementName>
   <type>...</type>
   <status date="...">...</status>
   <definition>...</definition>
   <context>...</context>
   <dataType>...</dataType>
   <representationalForm layout="..." minSize="..." maxSize="...">...</representationalForm>
   <dataDomain>...</dataDomain>
   <guideForUse>...</guideForUse>
   <verificationRules>...</verificationRules>
   <relatedDataReferences>...</relatedDataReferences>
   <sourceDocument>...</sourceDocument>
   <sourceOrganisation>...</sourceOrganisation>
   <comment>...</comment>
</dataElement>
```

One of the main advantages of using XML is that it is readable by both humans and machines. With practice, everyone can learn to read XML as if it were a book, and most modern software has XML capabilities built in. Web browsers such as Internet Explorer, for example, know that nested elements can be collapsed or expanded. This is indicated by the [+] and [-] buttons in Figures 15a and 15b below.

Figure 15a. XML document in Internet Explorer.

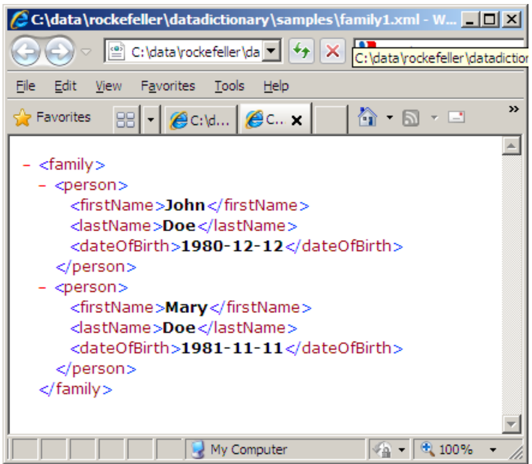
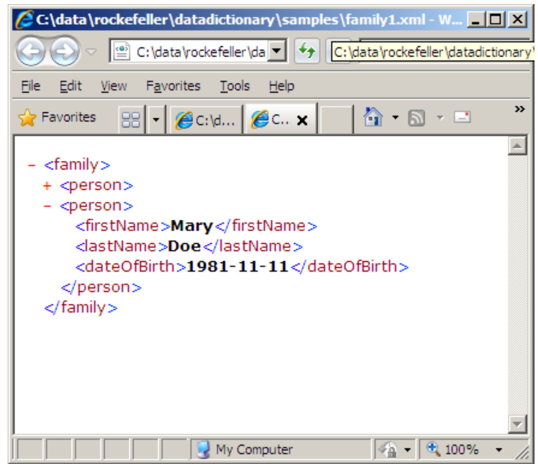


Figure 15b. John Doe “collapsed.”



Because XML is machine-readable, it is possible to have a software program look something up in an HDD without human intervention. Potentially, software could update its internal representations if the definitions in the central repository change.

To allow for such “programmatic” interaction with their national HDD, the Australians have written an application programming interface that can query and extract data from their METeOR website. See http://meteor.aihw.gov.au/content/index.phtml/itemId/430894.

e-claims

As suggested at the end of Part 1, it is often useful to start with those areas of an HDD that are core to the operations of a health insurance system. One such area is claims processing.

Between 2007 and 2010, the Dutch nongovernmental organization PharmAccess helped the state-run National Health Insurance Scheme in Ghana and the privately owned health insurer Strategis in Tanzania to create a common e-claims form based on information contained in their paper claims forms (Figures 16 and 17). By numbering the fields on the paper claims forms, mapping these numbered fields onto tables and columns in the respective databases, and then mapping these onto segments in electronic messages, both projects were able use the same e-claims format. The resulting XML structures are shown in Figures 18a and 18b.

The scanned paper forms in Figures 16 and 17 are not very legible because they had to be resized to fit on the page, but the contours of the various “data islands” should be clear enough to show that each item has been numbered.

Figure 16. Example claims form from Ghana.

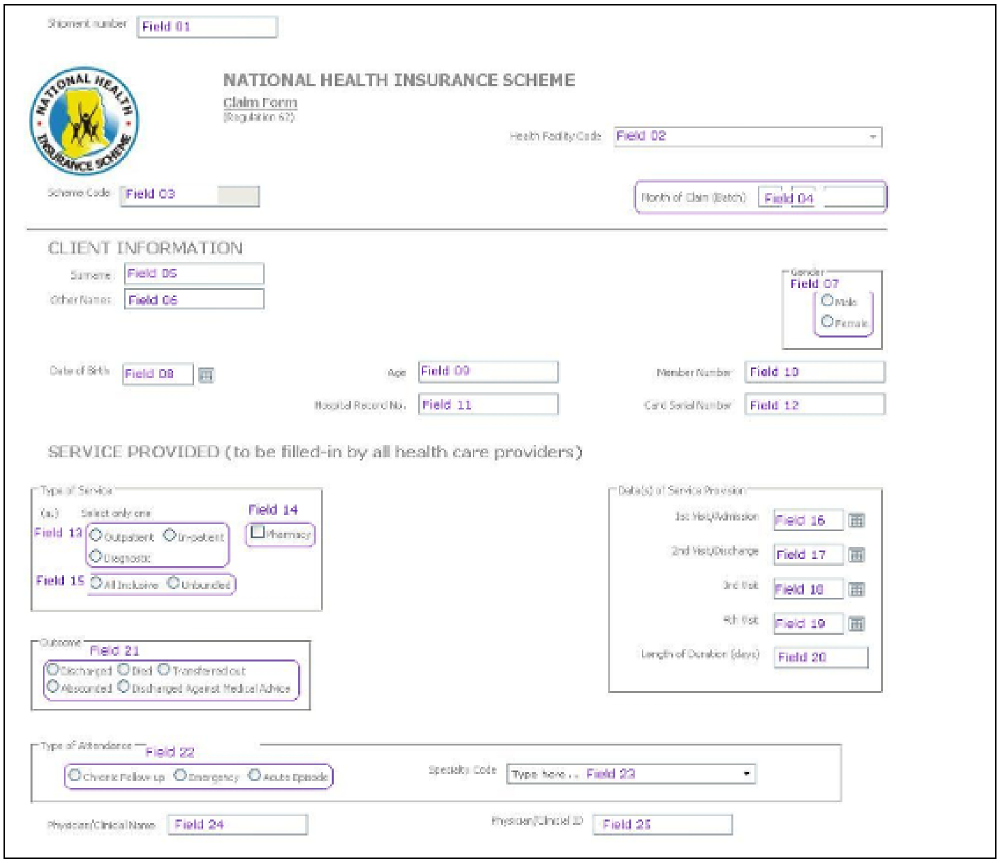

Shipment number Field 01
NATIONAL HEALTH INSURANCE SCHEME
Claim Form
(Regulation 62)
Health Facility Code Field 02
Scheme Code Field 03
Month of Claim (Batch) Field 04
CLIENT INFORMATION
Surname Field 05
Other Names Field 06
Gender Field 07 Male Female
Date of Birth Field 08
Age Field 09
Member Number Field 10
Hospital Record No. Field 11
Card Serial Number Field 12
SERVICE PROVIDED (to be filled-in by all health care providers)
Type of Service
(a.) Select only one
Field 13 Outpatient In-patient Diagnostic
Field 14 Pharmacy
Field 15 All Inclusive Unbundled
Date(s) of Service Provision
1st Visit/Admission Field 16
2nd Visit/Discharge Field 17
3rd Visit Field 18
4th Visit Field 19
Length of Duration (days) Field 20
Outcome Field 21
Discharged Died Transferred out
Absconded Discharged Against Medical Advice
Type of Attendance Field 22
Chronic Follow-up Emergency Acute Episode
Specialty Code Type here ... Field 23
Physician/Clinical Name Field 24
Physician/Clinical ID Field 25

Figure 17. Example claims form from Tanzania.

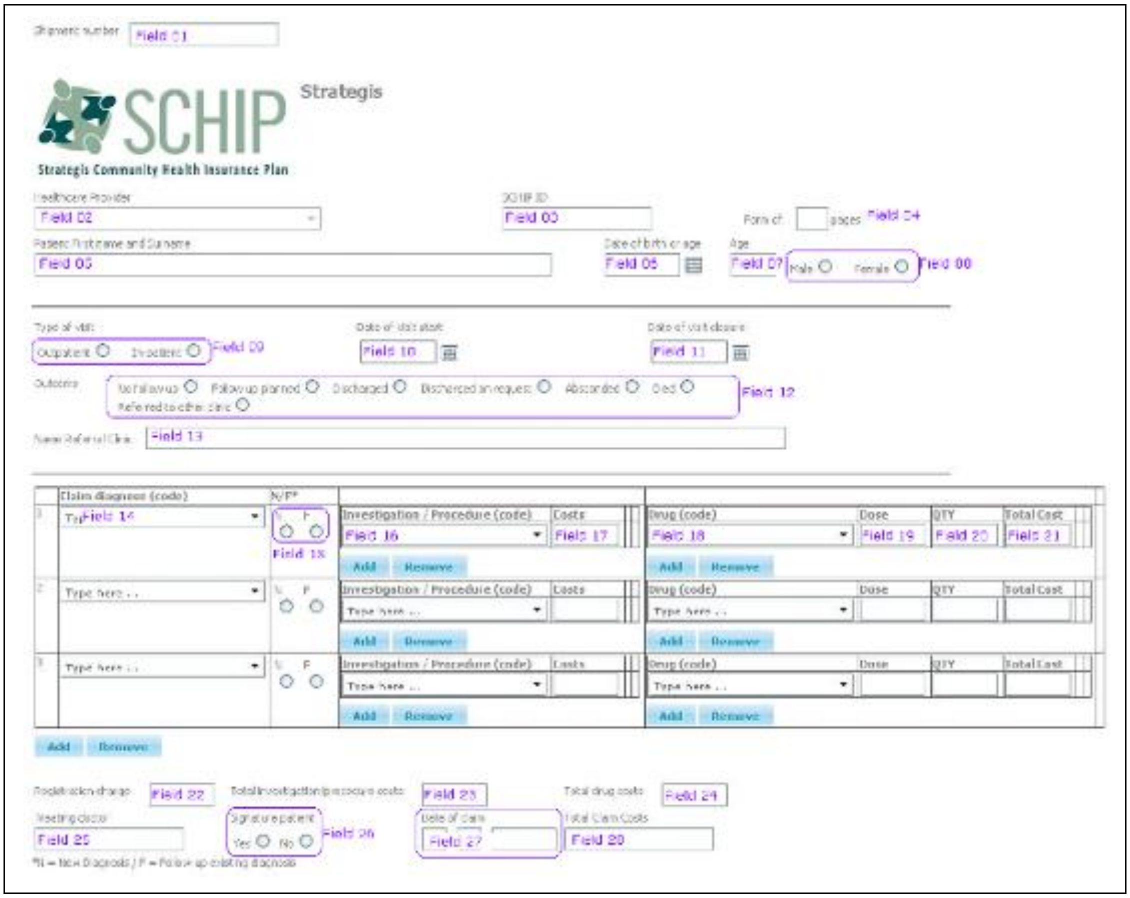


Figure 18a. Empty e-claim form in XML format.

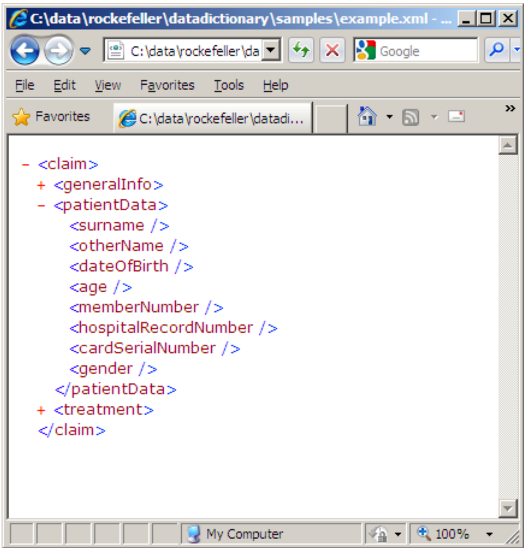
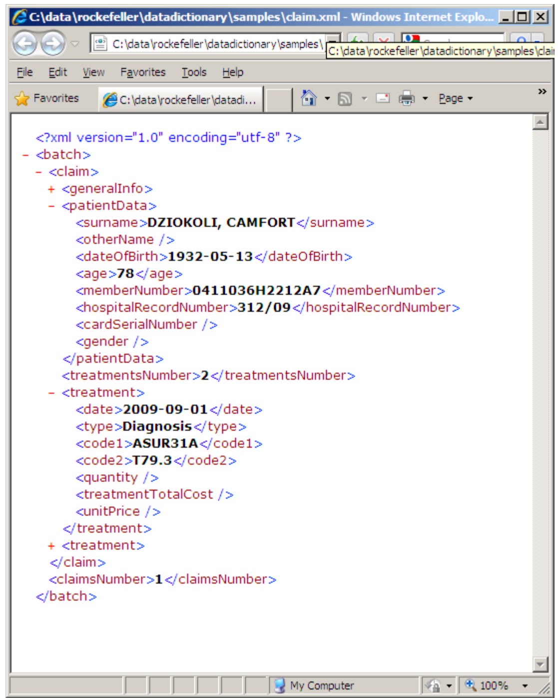


Figure 18b. Fictional completed e-claim form.

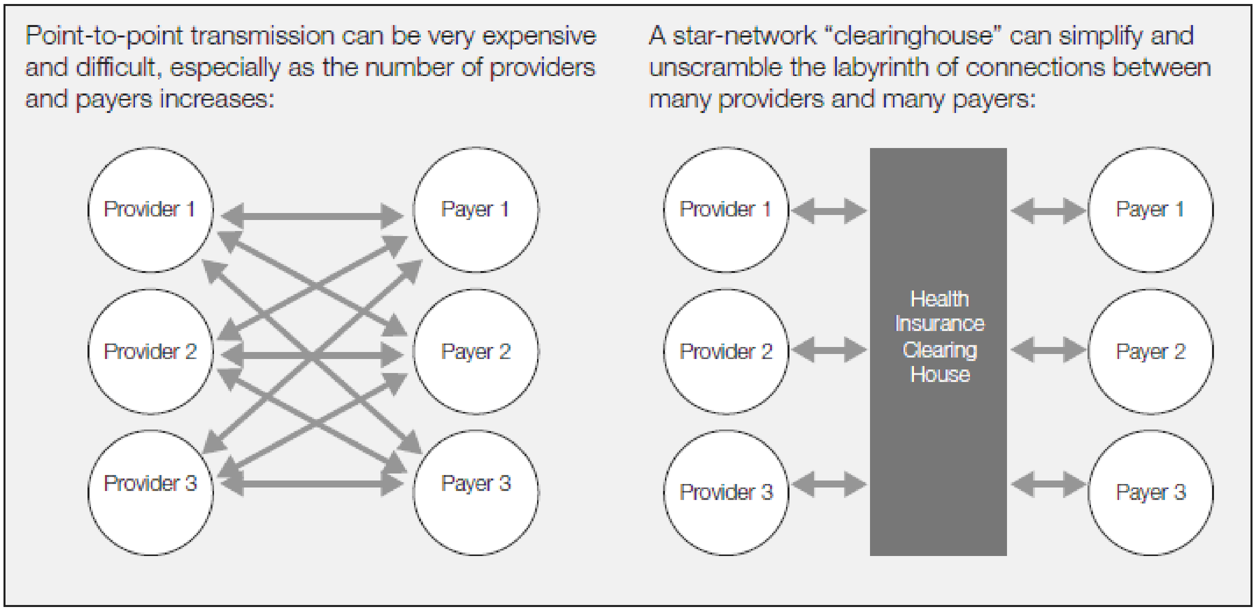


Architecture to the rescue

Without a national HDD, individual health care providers and payers must make their own arrangements on how to exchange information. Given that most countries have thousands of health care facilities—ranging from small pharmacies and private practices to community health centers and hospitals—and up to a few dozen payers, this information exchange becomes a labyrinth. Health care facilities dealing with different payers often must recode the information they collect for their own internal purposes, and then recode again when they submit data to local (district/regional) or national health authorities.

If all stakeholders in a country's health insurance system spoke the same "language" and adhered to the same standards and formats, more efficient information flows would be possible. A national HDD can serve as the basis for a hub-and-spoke or star-network model in which data need to be entered or transmitted only once. This is graphically depicted in Figure 19.

Figure 19. Point-to-point data transmission vs. a streamlined star-network model.



This idea was advocated by the World Bank and the US Agency for International Development in 2009 in an article called "Designing and implementing health care provider payment systems." How-to manuals are available at www.rbfhealth.org.

Claims routing in the Netherlands

The Netherlands started implementing this "hub" idea in 2005, both at the physical level (using a central claim-routing hub) and at the conceptual level (using a common set of standards). Over 99 percent of the 100,000,000 annual claims are now processed electronically, saving an estimated €400–600 million per year in administrative costs.

Figure 20. Key figures of claims handling in the Netherlands.

Key figures

- 100,000,000 — claims per year
- 40,000 — healthcare providers (professionals)
- 500 — claims software packages
- 10–20 — health insurers (payers)

Key solution characteristics

- 1 — Central claim-routing hub
- 1 — Common set of standards, message formats, interfaces
- 1 — National registry to identify providers, insurers, and patients respectively

Source: Ingun P, Streveler D, Brown K, et al. *The Role of Information Systems in Achieving Universal Health Coverage.* Geneva: WHO; 2010. Technical Brief Series—Brief No.10. Available at www.who.int/healthsystems/topics/financing/healthreport/ICTTBNo10.pdf .

HDD Prototyper Tool Guide

PART 3: HOW TO USE THIS TOOL TO HELP BUILD A NATIONAL FOUNDATION FOR INTEROPERABLE SYSTEMS

By Cees Hesp, Chief Technology Officer, PharmAccess Foundation

Introduction

Part 1 of this series provided an overview of how a health data dictionary (HDD) can promote interoperability, and Part 2 illustrated how different countries have tackled similar problems. Part 3 and the accompanying software is a practical tool for technical team members interested in replacing a paper dictionary with an electronic version from the outset.

Developed by PharmAccess Foundation, HDD Prototyper is a freely available tool that is designed to help countries develop an HDD. It allows you to write definitions for data items and then share this information as XML or HTML files. The XML file is intended for exchange with software vendors and other technical workers, and the HTML file can serve as documentation to be shared with other stakeholders. Both types of files can be shared on your organization's website or intranet for easy distribution.

This tool was developed under a grant from The Rockefeller Foundation (www.rockfound.org) to the Joint Learning Network Information Technology Track. The HDD Prototyper is freely available from www.jointlearningnetwork.org/content/tools for country use. Being a prototype itself, the tool is provided as-is, with no express or implied fitness for use. Please feel free to use and modify the tool as you see fit. Instructions for using the HDD Prototyper are provided below.

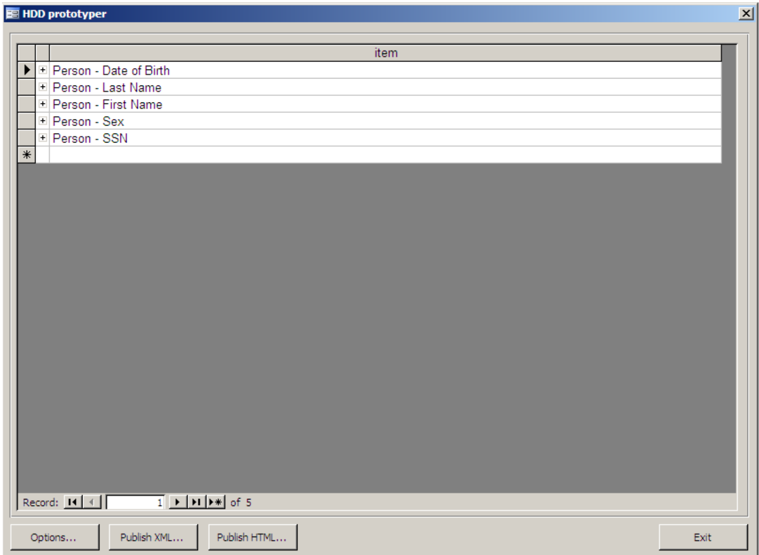
Getting started

HDD Prototyper is a Microsoft Access database distributed as a zip file. To work with it, you must first open the zip file and then extract the database to a location on your hard drive. Although you can open the database from within the zip file, it will remain in read-only mode and will not allow you to save any changes.

Start the application by double-clicking the "hdd.mdb" file. Depending on your computer's settings, the file might also display as "hdd" without the extension. Because Visual Basic for Applications was used to program the application, Access will probably warn you that the file is potentially unsafe. Make sure to scan the file for viruses, and then allow "macros" to run in Access. If you do not know how to do this yourself, ask a systems administrator to help you. Without macros enabled, the application will not work.

Once the file is running, Access should display a window or dialog box as shown in Figure 1. This is HDD Prototyper's main screen.

Figure 1. HDD Prototyper's main screen.

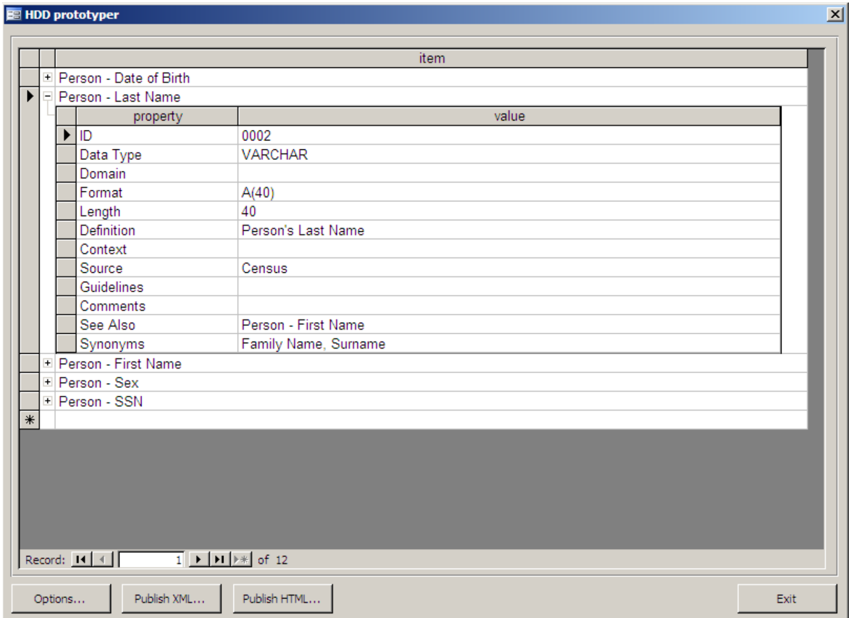


At the top of the screen, you will see a number of sample data items. At the bottom of the screen, you will see four buttons:

- [Options...] brings up a screen with different options to choose from.
- [Publish XML...] outputs your data dictionary to an XML file and then opens it in Internet Explorer.
- [Publish HTML...] outputs your data dictionary to an HTML file and then opens it in Internet Explorer.
- [Exit] asks for your confirmation to quit the application.

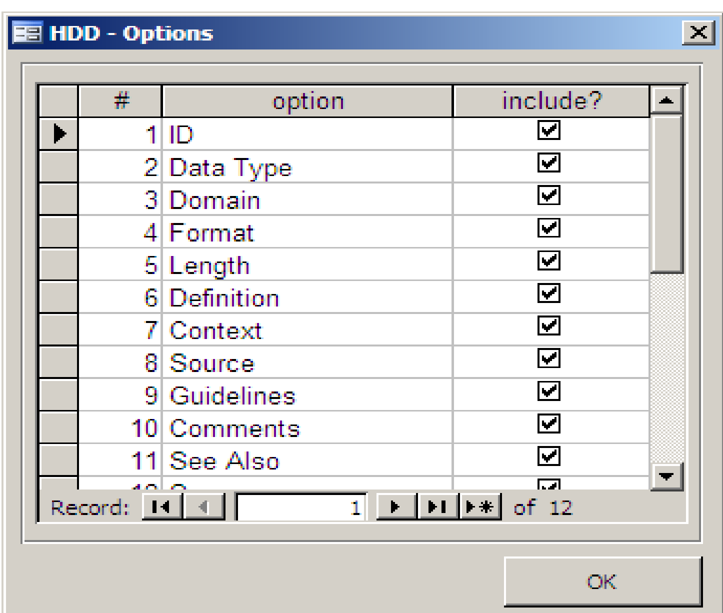
The items at the top are "collapsed" by default. You can expand each item to see details by clicking the [+] icon in front of it. The [+] icon then changes into [-]. Figure 2 shows what happens when you expand the "Person - Last Name" item.

Figure 2. HDD Prototyper's main screen with expanded details for "Person - Last Name."



When the details for all or some data items are expanded, the subdatasheet shows a number of properties (such as ID, data type, domain, and format) and their respective values. These properties can be used to describe data elements in some detail. If you want more or fewer of these properties, click the [Options...] button at the lower-left side of the screen. This will bring up a dialog box as shown in Figure 3.

Figure 3. HDD Prototyper's options screen.



There are several things you can do or change at the options screen:

- By selecting or clearing the checkmarks in the [include?] column, you determine which properties you want to use. For example, if you do not want to use the "Source" property, simply clear its checkmark.
- If you do not like the exact wording of a property, you can edit its name in the [option] column. For example, you could edit "Guidelines" to read "Guidelines for Use." You can do this for all of the options except "ID," "Length," and "See Also" (for reasons that will be explained later). Even though you cannot change the names of these three properties, you can choose not to use them by simply clearing their checkmarks.
- You can change the order in which the properties are displayed by manipulating the values in the [#] column. For example, if you wish to switch "Data Type" and "Domain," edit their sequence numbers to read 3 and 2, respectively. The sequence numbers can be anything you like as long as they are whole numbers (integers). Every time you change a sequence number, the screen will refresh itself to show the new order.
- If you scroll to the bottom of the list, you can add other properties as needed. This allows you to tailor the properties to your specific situation and needs. Remember that while you can add or change properties as much as you like, you cannot delete them. If you add a property and then choose not to use it, your only option is to clear its [include?] checkmark.

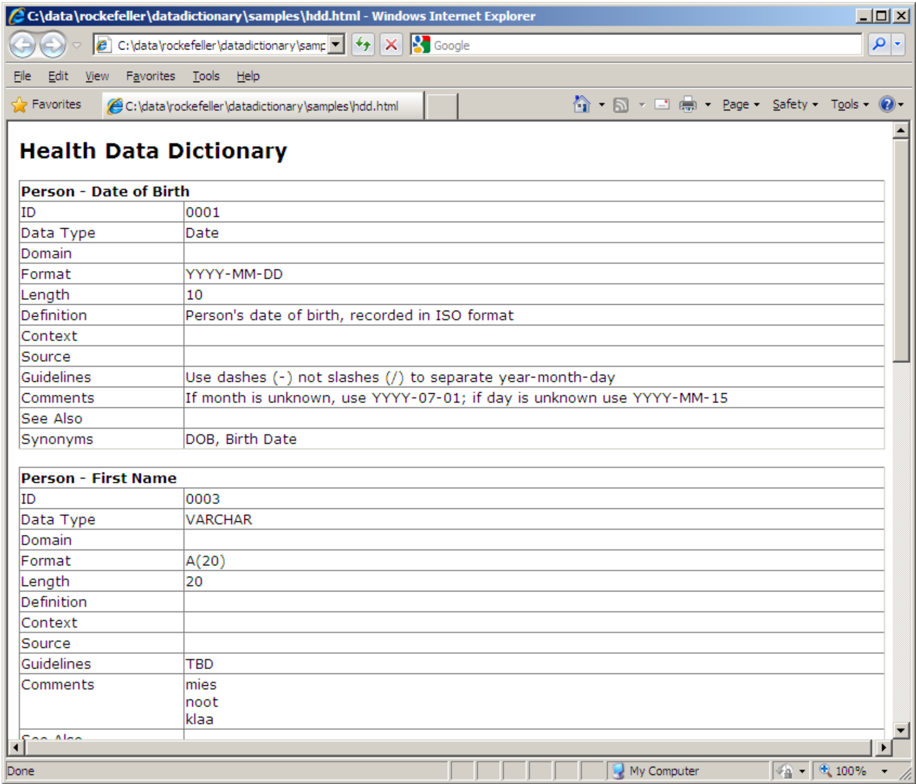
When you are done making changes, click [OK]. You can always return and make more changes.

After clicking [OK], all expanded subdatasheets with property details will collapse. When you reopen them, you should see the changes made in the options screen. If this does not work, close and then reopen the subdatasheets. And if Access does not automatically close the subdatasheets, closing and then reopening them will solve the problem.

Creating definitions

HDD Prototyper is a prototyping tool, meaning that you do not need to have all the details available before you start using it. Simply start entering the names of a number of data items and then click on one of the [Publish] buttons to see what the result looks like. Figure 4 shows an example of the resulting output.

Figure 4. Sample HTML (documentation) output.



Health Data Dictionary

Person - Date of Birth	
ID	0001
Data Type	Date
Domain	
Format	YYYY-MM-DD
Length	10
Definition	Person's date of birth, recorded in ISO format
Context	
Source	
Guidelines	Use dashes (-) not slashes (/) to separate year-month-day
Comments	If month is unknown, use YYYY-07-01; if day is unknown use YYYY-MM-15
See Also	
Synonyms	DOB, Birth Date

Person - First Name	
ID	0003
Data Type	VARCHAR
Domain	
Format	A(20)
Length	20
Definition	
Context	
Source	
Guidelines	TBD
Comments	mies noot klaa

Then start discussing which properties you want to record. Use the [Options...] button to tailor the system to your needs.

Once the data elements and their properties are more or less clear, start working on some real definitions. This is when you expand the subdatasheets and enter details. It is considered best practice for each data element definition to consist of some formal properties (e.g., Data Type, Format, Length, Definition) and some informational ones (e.g., Guidelines and Comments). The formal properties mostly address computer system requirements, and the informational properties aim to let human computer users know what to do with these definitions.

Some of the properties are fixed and cannot be changed, although you can choose not to use them. The properties with fixed names and meanings are:

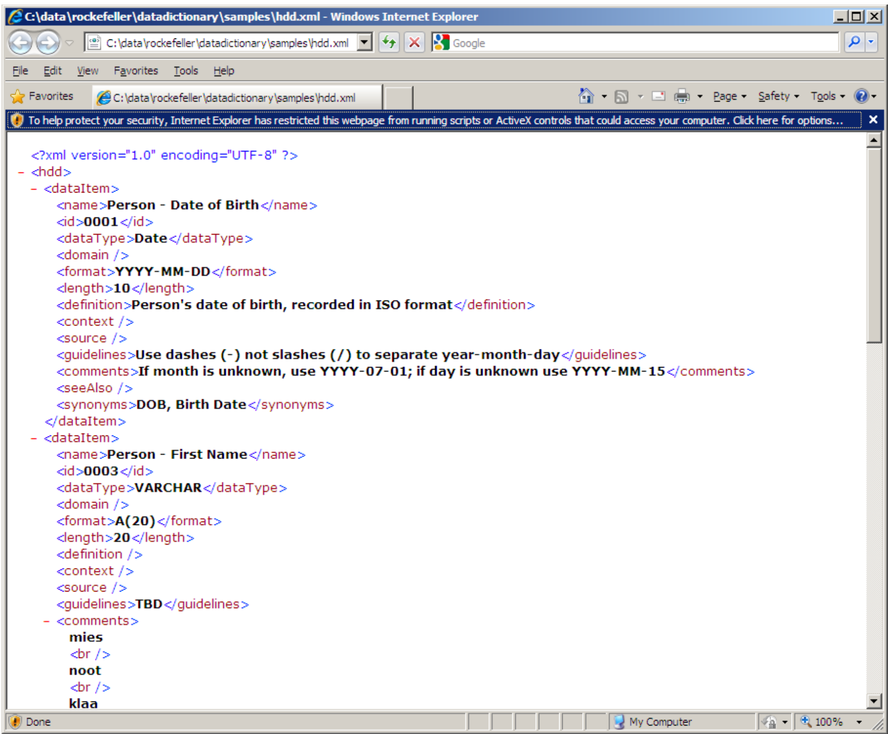
- **ID.** It is best practice to give each data element a unique ID. This is important because at some point, you will probably have hundreds or even thousands of data items with names that look similar. Giving each data item a unique reference ID helps to avoid confusion. In fact, HDD Prototyper requires that each ID be unique, regardless of whether you use plain numbers or a coding system.

- **Length.** When your definitions become requirements for software to be bought, built, or interacted with, the question of how much space is allotted to each data item will inevitably arise. If one system allows 50 characters for telephone numbers and another system only 10, these two systems may have a hard time communicating. Similarly, if your system requires 50 characters, systems only allowing 10 can easily be struck off the list. It is important, therefore, for the Length property to contain a proper number, and this is what HDD Prototyper enforces. You cannot enter "10.5" or "10_" (where the underscore represents a space). Neither of these would be values that a software system would know how to recognize for Length.
- **See Also**. A dictionary is not simply a list. You will need some mechanism to ensure cross-referencing. For example, the definition of "ABC" at the top of your dictionary may be related to definition "XYZ" at the bottom, with many pages in between. To make it easier to jump back and forth between definitions, the HTML documentation will automatically contain hyperlinks between related items. Clicking on such a link will take the reader directly to the item in question so there is no need to page through the document manually. For this reason, HDD Prototyper enforces that the value of the "See Also" property is the name of some other data item already defined. You do not want users to click on dead links. If a data item is related to more than one other item, please enter each reference (i.e., data item name) on a separate line.

Publishing XML

HDD Prototyper allows you to publish your HDD as an XML file. XML, short for eXtensible Markup Language, is a language that can be read both by humans and computers. Figure 5 shows an example.

Figure 5. Sample XML output.




```
<?xml version="1.0" encoding="UTF-8" ?>
- <hdd>
  - <dataItem>
      <name>Person - Date of Birth</name>
      <id>0001</id>
      <dataType>Date</dataType>
      <domain />
      <format>YYYY-MM-DD</format>
      <length>10</length>
      <definition>Person's date of birth, recorded in ISO format</definition>
      <context />
      <source />
      <guidelines>Use dashes (-) not slashes (/) to separate year-month-day</guidelines>
      <comments>If month is unknown, use YYYY-07-01; if day is unknown use YYYY-MM-15</comments>
      <seeAlso />
      <synonyms>DOB, Birth Date</synonyms>
    </dataItem>
  - <dataItem>
      <name>Person - First Name</name>
      <id>0003</id>
      <dataType>VARCHAR</dataType>
      <domain />
      <format>A(20)</format>
      <length>20</length>
      <definition />
      <context />
      <source />
      <guidelines>TBD</guidelines>
    - <comments>
        mies
        <br />
        noot
        <br />
        klaa
```


The output shown in Figure 5 indicates that there is something called an “<hdd>” (i.e., health data dictionary). This HDD consists of a number of elements called “<dataItem>,” two of which are visible in Figure 5. The data items themselves are made up of elements called “<name>,” “<id>,” “<dataType>,” etc. These terms in angle brackets are *exactly* what you defined as items and properties in your data dictionary. Now you have written your dictionary in such a way that the definitions can be read by computers. A computer can actually read the document shown in Figure 5 and extract the fact that a person’s date of birth must be expressed as year-month-day, for example, and that the representation has a maximum length of 10 characters. If this definition were to change over time, such as when four digits no longer suffice to represent years, the computer system could conceivably update its own internal date representation to comply with your new definition.

The example above (about changes over time in the number of characters needed to indicate year of birth) may seem far-fetched because years can be represented as four digits for nearly 8,000 years to come. Remember, though, that this work is shaping your country’s future. In that distant future, the decisions you make today can affect your country in unexpected ways. For example, if your country has a rapid increase in the use of mobile phones, the number of digits needed for that definition may change. Similarly, dramatic changes in the value of your country’s currency might require changes in related definitions.

This does not mean that you have to foresee and plan now for every possible change. It means that you need a mechanism to communicate changes. And that is where XML and similar technologies (e.g., JSON, or JavaScript Object Notation) can play an important role. These languages are designed to communicate about the structure and contents of electronic documents.