3M Health Information Systems

White Paper

Achieving Semantic Interoperability

March 29, 2006

Executive Summary

As defined by the National Committee on Vital and Health Statistics (NCVHS) in their July 6, 2000 *Report on Uniform Data Standards for Patient Medical Record Information*, interoperability refers to the ability of one computer system to exchange data with another computer system. There are three levels of interoperability, each adding to the previous layer:

- **"Basic" Interoperability**
  - Messages be can exchanged between systems, but not interpreted.
- **"Functional" (Syntactic) Interoperability**
  - The structure, or format, of the message is defined (hence the term message or format standard), but the meaning of the data within the data fields is not understood by either systems.
- **"Semantic" Interoperability**
  - The meaning of the data within the data fields of the common message structure is understood by both systems.

Efforts towards interoperability have accelerated over the last few years. Standard developing organizations such as Health Level 7 (HL7) have taken the lead in providing interoperability definitions, and the HL7 version 2.x messaging standard is implemented in information systems worldwide. Most, if not all, clinical information systems are currently compliant to messaging standards, making syntactic interoperability a reality.

However, semantic interoperability is still uncommon, for many reasons. There are many terminologies being used in health care, because no single medical terminology covers all operational needs. Nevertheless, the foundation has been laid and overall, good progress in terminology development has been made and standards are available to support semantic interoperability. The challenge remains that of implementation.

Most EHR development predates current semantic interoperability requirements, and most health care organizations already have clinical information systems in place, a significant financial as well as operational investment. Replacing these legacy EHRs is not be a viable option for many organizations. Considerable patient data has already been collected, encoded, and stored with the legacy terminologies in these information systems. This historical data is invaluable clinical information that is critical to continuity of care and optimal outcome for the patient. Switching to encoding new patient data with a different terminology, even a standard, would mean these historical data is no longer compatible to the new data. Thus, while the new data encoded with a standard terminology is semantically interoperable with other external data also coded to the same standard, ironically, the organization's own historical data would not be.

A practical migration path is needed to help health care organizations gain the ability to achieve semantic interoperability, without imposing undue burden or resulting in the loss of historical patient data. 3M HIS has developed its Care Innovation EHR products to work with other clinical information systems in a best-of-breed approach, recognizing that health care organizations may need to preserve their current applications and data. The 3M Enterprise Master Patient Index (EMPI) and Clinical Data Repository (CDR) provides for accurate patient identification and structured, encoded data storage across the entire health care enterprise. The 3M Healthcare Data Dictionary (HDD) provides a master reference terminology used to encode data, from incoming messages, for storage in the CDR.

The HDD is cross-referenced to standard terminologies. If the sending system uses non-standard codes in their outbound messages, then these legacy codes are first mapped into the HDD. Through mapping, the HDD translates between standard terminologies, between legacy systems, and between a legacy system and a standard terminology. For external data exchange, the HDD translates the encoded data to the required standard code. Thus, using the HDD, semantic interoperability is achieved with no need for health care organizations to change the system-specific terminologies that they are currently using to support their workflow.

The Care Innovation approach has been implemented in commercial health care enterprises and the Department of Defense (DoD) Armed Forces Health Longitudinal Technology Application (AHLTA) project, the latter being deployed at all Military Treatment Facilities (MTFs) worldwide. The 3M solution proves a practical approach to achieving data interoperability that can serve as a case study for the nation.

Interoperability

What is interoperability?

The goal of interoperability is to be able to do more with data than merely capture, store and display it in a single, standalone system. Advanced levels of interoperability empower decision makers with actionable information by enabling their computer systems to exchange meaningful data and to interpret and act upon shared data and knowledge. As defined by the National Committee on Vital and Health Statistics (NCVHS) in their July 6, 2000 *Report on Uniform Data Standards for Patient Medical Record Information,* there are three levels of requirements:

- **"Basic" Interoperability**
  - Messages be can exchanged between systems, but not interpreted.
- **"Functional" (Syntactic) Interoperability**
  - Messages can be exchanged between systems and interpreted, but only to the level of the data fields; that is, the structure, or format, of the message is defined (hence the term message or format standard), but the meaning of the data within the data fields is not understood by either systems.
- **"Semantic" Interoperability**
  - Messages can be exchanged between systems and interpreted by both systems; the structure of the message is defined and the meaning of the data within the data fields is understood. The data can be acted on by the receiving system automatically.

Take this example: A severely ill infant with suspicion of meningitis is transferred from a rural medical facility to a pediatric hospital. The health information systems at the two facilities are not interoperable and cannot share information regarding the patient. At the rural facility a Cerebral Spinal Fluid (CSF) sample was collected and laboratory tests performed, but when the baby arrives at the pediatric hospital, none of the previous results from the rural hospital are available to the new physician. The parents insist that testing was performed and may even know, for example, that bacterial meningitis was ruled out. But it does not matter, the new physician must start over and this begins with the collection of a new CSF specimen.

Collecting a CSF sample on an infant can be a traumatic procedure for both the clinician and the child and can have complications that worsen the outcome. Not only does the duplicate testing introduce additional risk, pain, and expense, but also causes a delay in this time critical, life threatening situation. The example above describes systems that have no interoperability. The following describes information exchange through progressively higher levels of interoperability:

- **"Basic" Interoperability**
  - The physician must search through printouts or multiple screens of data to see if the necessary laboratory test results are available.

- **“Functional” Interoperability**
  - The receiving system is able to recognize a laboratory test result (versus a medication order, for instance) in another system, but results for a particular test cannot be identified and the values cannot be interpreted or aggregated with results in the receiving system.
  - Automated alerting for duplicate testing, adverse events or infectious diseases will not work for data collected from other systems.
- **“Semantic” Interoperability**
  - The receiving system is able to recognize and interpret a particular laboratory test and its results from another system.
  - Critical information vital to accurate decision making is readily available to automated decision support and the clinician without additional time, expense, or hardship for the patient.

Interoperability also has significant implications at the population level. Disease surveillance and early alerting depends heavily on the timely and accurate exchange of clinical information. Without basic interoperability, we would need humans manning phones and fax machines. With syntactic interoperability, data from across the nation can be collected in a single database without needing manual data entry, but a human would need to read and interpret the data to spot the disease occurrences. With semantic interoperability, the system can recognize the disease occurrences without human intervention, and automatic triggers could be set up for alerts.

Interoperability: Toppling the Tower of Babel in Health Care

Empower decision makers with actionable information by enabling their computer systems to exchange meaningful data and to interpret and act upon shared data and knowledge.

<u>What does interoperability mean to health care delivery?</u>

Fundamentally, interoperability relates to the need to communicate data between systems. Since communication of data is essential to achieve the best efficiency and outcomes in health care, interoperability is critical. Without interoperability, we can carry on delivering care – and have – but there are waste and adverse effects, and we cannot take maximum advantage of our systems and data. The examples in the section above illustrate the duplicated clinical effort (e.g., laboratory testing), additional human intervention (e.g., manual data entry and review), and difficulty in performing advanced data functions (e.g., automatic alerts) without interoperability.

The following attachment is a literature review evaluating the potential return on investment (ROI) of clinical information systems in general, and highlights the impact from a lack of interoperability.

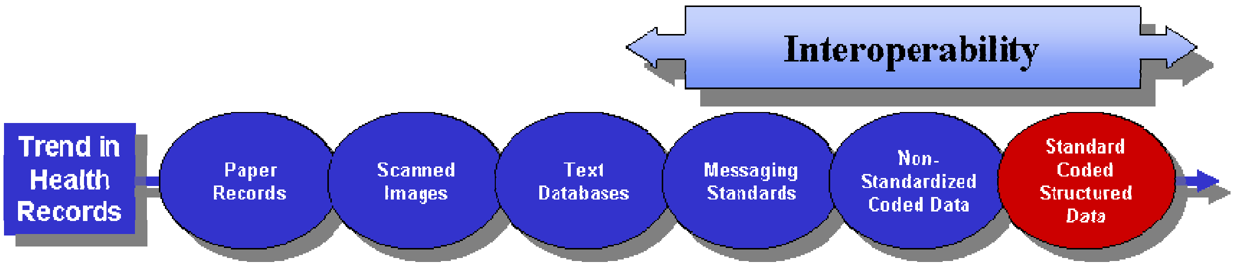
CIS ROI

The state of Maine commissioned a cost-benefit analysis as part of the feasibility study for their clinical information sharing project, the Maine Health Information Network Technology (MHINT) system. The cost-benefit analysis concluded that a clinical information sharing system could potentially save the state of Maine $42.3-$58.4 million annually in healthcare costs. The cost-benefit analysis is included in the MHINT *Phase I Report*, which can be found on the MHINT website (www.mhint.org). In their January, 2005, *Health Affairs* article, "The Value of Health Care Information Exchange and Interoperability," Walker *et al* from the Center for Information Technology Leadership in Boston, MA, assessed the value of electronic health information exchange. They concluded that fully standardized health information exchange and interoperability could potentially result in a net value to the U.S. of $77.8 billion annually. The article is attached below.

Value of HIEI

The Electronic Health Record and Interoperability

Progression in computerization of health care data



We have long recognized that a paper health record has multiple problems from a data sharing standpoint. The immediately obvious ones are due to the physical nature of the paper record:

- Not sharable unless duplicated; then the multiple copies will go out of synchronization when new data is added independently to each.
- Manual effort is required to find the record when it is needed.
- Susceptible to loss and damage.

These problems can be overcome by the most primitive capabilities of computerization – scanning the paper record into electronic form. However, additional problems with regard to the usability of the data remain:

- Data can be unreadable or misinterpreted due to handwriting issues, low ink, poor scanning, etc.
- Data cannot be edited or updated – because scanned images are essentially "blobs" in electronic form.

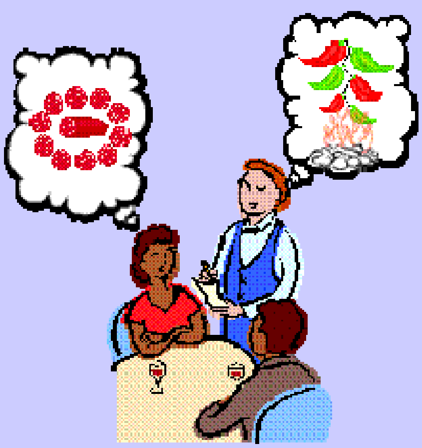
Capturing data as electronic text (free or narrative text) can overcome some challenges with regard to readability (setting aside issues of poor spelling or grammar, etc.), but is insufficient for interoperability, because human language is so variable. For instance, "Symptom: Headache; Diagnosis: Hypertension" and "Diagnosis: Hypertension; Symptom: Headache" are two entirely different strings of text to a computer. Thus, the next step towards interoperability is to define a common message structure – for example, data field 1 for symptom and data field 2 for diagnosis – generating a standard for functional (syntactic) interoperability.

- What about the data in the data fields within the message? In theory, free text could be used to achieve a common understanding of data between different systems, provided that the exact same text is always used to describe the same thing. For instance, a clinician would recognize "high blood pressure", "hypertension", "HPT" and "HBP" as the same diagnosis in a patient's record. However, to a computer, these are different entities unless it is specifically instructed that they are merely different representations of the same concept. In reality, human beings are so creatively variable that once it is possible to collect large amount of data electronically, the countless variations become quickly apparent. Data collected as electronic free text requires a large amount of additional effort to normalize and interpret.

Correct identification of the data is required to provide decision support in care delivery. For instance, in order to alert the physician ordering medication for a patient to a possible allergy, e.g. penicillin, all penicillins must be correctly identified regardless of their name, e.g., amoxicillin, ampicillin, “Pen VK”, etc. Free text is also very difficult to compare and aggregate for analysis. Because of these reasons, clinical information systems have taken to encoding data instead of storing free text. This means a single code is used to store a data item, for instance, “12345” instead of “high blood pressure”, “hypertension”, “HPT”, “HBP”, and all other variations. Human variability inherent in language is thus removed by data encoding.

One last challenge remains. For many good reasons, different computer systems may use different sets of codes for data encoding. Within each system, data encoding may be consistent, but different systems are unable to understand one another’s codes. Meaningful data cannot be exchanged or connected to external knowledge sources. In an attempt to provide a means to a common understanding of the data, standard terminologies are developed. If all systems use the same messaging standards and the same standard codes to communicate data, then the data becomes understandable to all. A system may choose to use its own proprietary code sets to encode the data, but would translate to the standard code to send the data externally or reference outside knowledge sources. This is analogous to everyone speaking the same, selected, language at an international conference, regardless of what language is spoken at home.

Semantic vs. Syntactic Interoperability



For an everyday analogy, imagine a German waiter and an American diner attempting to communicate in English.

The American says, “I have eaten pepperoni pizza before, so I think I will try this salad.” The German may recognize all the terms used by the American, but information critical to successful communication may not be shared.

The required elements are a shared grammar (functional interoperability) to understand that the above statement is in past perfect, “I have eaten”, not “I would like to eat”; and a shared vocabulary (semantic interoperability). In British English, pepperoni is jalapeño pepper.

The consequence of sharing terms but failing to reach functional and semantic interoperability is that the American diner is about to receive a jalapeño pizza and a salad for dinner.

What is a standard terminology?

A standard terminology is one that has wide industry acceptance or use. Standards are obtained from a variety of efforts. The federal government has purchased a national license for the Systematized NOmenclature of MEDicine Clinical Terms (SNOMED CT). The National Library of Medicine (NLM) maintains the Unified Medical Language System (UMLS), with the latest focus being RxNorm, a reference terminology for clinical drugs. Standards are developed by consensus industry effort, such as HL7 version 3. The set of standards to be used by government agencies will be named by the Consolidated Health Informatics (CHI) initiative – the first terminology standard selected is Logical Observation Name Identifiers and Codes (LOINC). Examples of terminologies that are considered standards for billing are the International Classification of Diseases, 9th Edition, Clinical Modification (ICD9CM) and the Common Procedural Terminology (CPT). The National Drug Code (NDC) may also be considered a standard for use within the pharmacy industry.

The current state of EHR and interoperability

Efforts towards interoperability have accelerated over the last few years. Standard developing organizations such as Health Level 7 (HL7) have taken the lead in providing interoperability definitions, and the HL7 version 2.x messaging standard is implemented in information systems worldwide. Most, if not all, clinical information systems are currently compliant to messaging standards, making syntactic interoperability a reality. However, semantic interoperability is still uncommon, for many reasons. There are many terminologies being used in health care, because no single medical terminology covers all operational needs:

- Different standard terminologies are developed for different health care domains (e.g., laboratory, pharmacy, etc.)
- Different standard terminologies are developed for different purposes (e.g., reimbursement, clinical documentation, etc.)
- Different health care applications developed different system-specific terminologies
- Health care facilities have developed their own enterprise-specific terminologies in addition to, or instead of, standards

Nevertheless, the foundation has been laid and overall, good progress in terminology development has been made and standards are available to support semantic interoperability. The challenge remains that of implementation. Most EHR development predates current semantic interoperability requirements, and most health care organizations already have clinical information systems in place, a significant financial as well as operational investment. Replacing these legacy EHRs is not be a viable option for many organizations. Considerable patient data has already been collected, encoded, and stored, using the legacy terminologies in these information systems. This historical data is invaluable clinical information that is critical to continuity of care and optimal outcome for the patient. Switching to encoding new patient data with a different terminology, even a standard, would mean these historical data is no longer compatible to the new data. Thus, while the new data encoded with a standard terminology is semantically interoperable with other external data also coded to the same standard, ironically, the organization's own historical data would not be.

A practical migration path is needed to help health care organizations gain the ability to achieve semantic interoperability, without imposing undue burden or resulting in the loss of historical patient data. 3M HIS has developed its Care Innovation EHR products with this goal in mind. The Care Innovation approach has been proven in commercial health care enterprises and the Department of Defense (DoD) Armed Forces Health Longitudinal Technology Application (AHLTA) project, the latter being implemented at all Military Treatment Facilities (MTFs) worldwide. The 3M solution proves a practical approach to achieving data interoperability that can serve as a case study for the nation.

A Practical Approach to Interoperability

Challenges in implementing semantic interoperability

As we have discussed, health care organizations are beginning to recognize the many advantages of semantic interoperability. However, achieving semantic interoperability presents some challenges. Standard terminologies are a means of encoding data for exchange, comparison or aggregation among systems. However, there are issues concerning their practical use in the EHR. These include:

Semantic shift: shift in the meaning of a standard code. When a standard code is used to encode data directly in the EHR and the meaning of the code changes over time (for example, due to code reuse), historical patient data will be interpreted incorrectly.

Code deprecation: deletion of a standard code. When data is encoded with a standard code that has since been removed by the Standard Developing Organization (SDO), the data is no longer interpretable.

Lack of comprehensive standard codes: A standard vocabulary may not provide all the codes that are needed to encode the entire set of data in current use.

Local extensions: codes added at each facility to support operational needs, e.g., locally formulated medications.

Historical patient data: If historical data has been encoded with non-standard terminologies, then, encoding new data using standard terminologies from this point onwards would result in the historical data not being interoperable with the new data.

The 3M approach to interoperability

The 3M Healthcare Data Dictionary (HDD) integrates and supplements standard vocabularies, presenting a master reference terminology used to encode data in the Clinical Data Repository (CDR). The HDD content is cross-referenced to standard terminologies. Each HDD concept is identified by a Numerical Concept IDentifier (NCID), and the identifiers from external terminologies are mapped to it. NCIDs are used to encode the data in the CDR. The code in the incoming message to the CDR is translated by the HDD to the corresponding NCID for storage. If the sending system uses non-standard codes, then these legacy codes are first mapped into the HDD. Through mapping, the HDD translates between standard vocabularies, between legacy systems, and between a legacy system and a standard vocabulary. For external data exchange, the HDD translates the encoded data from NCIDs to a standard code. Thus, using the HDD, semantic interoperability is achieved with no need for health care organizations to change the system-specific terminologies that they are currently using to support their workflow.

The 3M HDD addresses the challenges of implementing semantic interoperability in the following manner:

Semantic shift: overcome because HDD concepts never change their meanings, thus the reused code will be mapped to a different concept, and data encoded with NCIDs will not be misinterpreted.

Code deprecation: overcome because an HDD NCID is never deleted; an inactive standard code is marked as such without affecting the CDR data.

Lack of comprehensive standard codes: overcome because the HDD provides all concepts needed to encode data in the CDR. If a new code is later assigned by the standard vocabulary, it is mapped to its corresponding NCID in the HDD, with no updates required for the data stored in CDR.

Local extensions: interoperable because the HDD coordinates all local extensions within an organization so the data is standardized enterprise-wide, and mapped to standard terminologies.

Historical patient data: interoperable because the HDD maps legacy codes to NCIDs to ensure backward compatibility.

3M has designed its Care Innovation products to work with other clinical information systems in a best-of-breed approach, recognizing that health care organizations may need to preserve their current applications and historical data. Working with the HDD, the Enterprise Master Patient Index (EMPI) and CDR provides for accurate patient identification and structured, encoded data storage across the entire health care enterprise.

3M Care Innovation solutions

Clinical Data Repository (CDR). The 3M CDR is a central database of individual, electronic, lifetime patient records that users can access and analyze and to which they can add information, right at the point of care. The 3M CDR supports HL7 standards. While data in the CDR comes primarily from interfaced legacy information systems, clinicians can also document findings directly into the CDR. Once stored in the CDR, the complete longitudinal patient record immediately becomes available to authorized users throughout the healthcare continuum, such as acute care facilities, clinics, surgery centers, wellness and home healthcare centers, and doctors' offices. The 3M CDR is unique in its robust information models and data architecture, ensuring that patient data is stored accurately and consistently in a strongly structured and encoded form. This allows the data to be used in a longitudinal as well as population-based manner, enabling alerts and other decision support, outcomes research, and quality improvement. In conjunction with the Healthcare Data Dictionary (HDD) and the Enterprise Master Person Index (EMPI), the CDR enables true data interoperability by storing and presenting the data in a consistent, structured, encoded format, and in compliance with HL7 standards.

The attached CDR fact sheet provides additional information.

3M CDR

3M Enterprise Master Person Index (EMPI). 3M's EMPI lets an enterprise quickly and consistently identify a patient across all care sites and healthcare encounters. The EMPI integrates and consolidates patient indices from multiple registration systems into a single index, providing an accurate, centralized summary of both patient demographic and encounter information. The EMPI tracks data for the following:

- Patients who have seen a physician or have been admitted to a hospital
- Members of enterprise health plans and wellness facilities
- Subscribers with no healthcare encounters
- Guarantors and other entities

Regardless of where an individual receives care, the EMPI saves registration and admissions personnel time at check-in by giving them access to that person's most current data. And, as patients move between care sites without having to repeat their admissions data, they perceive the enterprise as a single organization that recognizes them as individuals and understands their healthcare needs.

The attached EMPI fact sheet provides additional information.

3M EMPI

3M Healthcare Data Dictionary (HDD). The 3M HDD is a database that describes the organization and logical structure of the medical data found in the CDR. It contains "metadata" – "data about data" – that describes the content and structure of, and relationships between, clinical data in the CDR. The HDD is a concept-based terminology server that provides terminologies used in health care, integrated in a seamless manner. It is a medical data dictionary that encodes clinically relevant patient data and is targeted at mapping among standard and interface terminologies, unifying them under a single, consistent model. In short, the HDD "translates," precisely defines, and effectively accesses the contents of the EHR. Data enters, is stored, and leaves the 3M CDR according to the encoding information provided by the HDD.

The attached HDD fact sheet provides additional information.

3M HDD

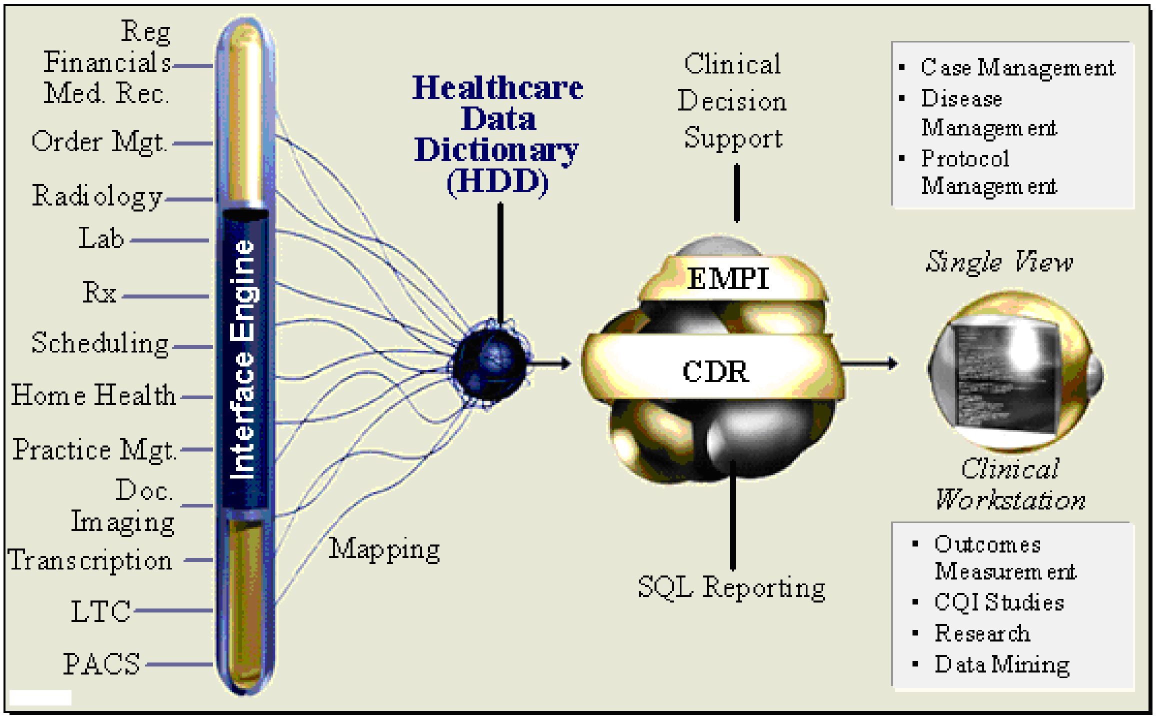
The Armed Forces Health Longitudinal Technology Application (AHLTA)

The Department of Defense (DoD) is currently leading health information technology initiatives with the implementation of AHLTA, the DoD's web-based EHR that, upon full deployment to all DoD Military Treatment Facilities (MTFs) worldwide, could be the largest structured medical data system in the world. AHLTA currently supports over 6.2 million patients, including 1.4 million active duty military personnel, their families, and a number of military retirees. The project's budget is $3-4 billion over an 18-year life cycle. The system provides authorized users with secure, electronic access to a patient's comprehensive health record and includes data on preventive care, illness, and injuries. When fully deployed, AHLTA will be used by over 30,000 providers worldwide, including physicians, nurses, and ancillary staff.

3M HIS is an integral part of this DoD effort, which implements the 3M Care Innovation products, including the 3M Healthcare Data Dictionary (HDD), Clinical Data Repository (CDR), and Enterprise Master Person Index (EMPI), in AHLTA. This ambitious effort builds upon a previously deployed EHR in the DoD – the Composite Health Care System (CHCS), AHLTA's predecessor – and enhances that foundation through the provision of the 3M solution for interoperability.

In the late 1980s and early 1990s, the DoD deployed CHCS across all MTFs. While the medical information is computerized, each CHCS host system (there are over 100) functions as a standalone EHR, with each having its own system specific terminology that has been used to encode data. Thus, the CHCS data is not semantically interoperable from one DoD MTF to another. With the 3M approach to semantic interoperability, the DoD is able to continue CHCS operation at each MTF, while deploying the 3M CDR, EMPI and HDD as the enterprise-wide EHR, providing a transition that preserves its investment in CHCS and avoids interruption to operation and workflow. In January 2004, the DoD began the worldwide deployment of the new AHLTA system. System specific terminologies from all CHCS host systems have been mapped into the HDD, which is used to encode data from CHCS into the CDR in AHLTA. The data from all DoD MTFs are thus interoperable across the entire DoD. Historical patient data has been preloaded into the CDR, also encoded with HDD terminology, and is thus interoperable with new data. Through the HDD, the DoD data can be translated to the required standard code for semantic interoperability with external organizations such as the Veterans Administration.

The following figure 1 illustrates how 3M's CDR, EMPI, and HDD have been implemented in AHLTA.



Data from over 60 clinical applications is interfaced into AHLTA, forming a longitudinal record of an individual, and then the data is made available to the hundreds of other treatment locations that the individual may visit. Typical weekly statistics for the system include the following:

- > 15,000 admissions
- 1.7 million outpatient encounters
- 1.9 million prescriptions
- > 1,800 births
- 400,000 dental encounters
- > 580,000 dental procedures

Currently AHLTA is over 50% implemented, and system installation is projected to be complete by December 2006. At full deployment AHLTA will support 9.2 million beneficiaries, including families and retirees.

Conclusion

Semantic interoperability defines a level of communication between disparate computer systems at which both the "grammar" (message structure) and "vocabulary" (terminology/codes used to store data) of messages used for data exchange are shared and explicitly defined. Even though it is critical to enhancing the efficiency and quality of health care delivery and outcomes, most health care information systems have not yet achieved a useful level of interoperability. This failure can be ascribed to many factors, among them: legacy clinical systems and data are valuable and expensive to replace or redefine and while excellent standards for messaging and clinical terminology are available, a migration path is not clearly seen. A practical implementation and maintenance strategy is needed.

3M Health Information Systems (HIS) has developed Electronic Health Record (EHR) products and services to address these challenges and help organizations achieve interoperability and maximize the usefulness of their data. The 3M Clinical Data Repository (CDR), Enterprise Master Patient Index (EMPI) and Healthcare Data Dictionary (HDD) are used in commercial health care enterprises and the Department of Defense (DoD) Armed Forces Health Longitudinal Technology Application (AHLTA) project, the latter being implemented at all Military Treatment Facilities (MTFs) worldwide. The 3M solution proves a practical approach to achieving data interoperability that can serve as a case study for the nation.