Terminology Services Bureau

February 18, 2004

EXECUTIVE SUMMARY

Introduction

To support data interoperability, the Department of Defense (DoD) needs to manage the implementation and use of standard coding schemes (e.g. Common Procedural Terminology – CPT, and Logical Observation Name Identifiers and Codes – LOINC), source vocabularies (e.g. Medcin and Problem Knowledge Coupler – PKC), and interface terminologies (e.g. Composite Health Care System – CHCS master files) as an integrated solution.

Accomplishments

Clinical Care

The DoD has made a commitment to data interoperability through standardization in the 3M Healthcare Data Dictionary (HDD). The CHCSII data standardization accomplished by the DoD is unprecedented in scale and implication. The DoD will be able to share accurately and consistently coded data for patient care among Military Treatment Facilities (MTFs), and also enable accurate, efficient and useful decision support, patient safety efforts and population based applications. The HDD will translate the DoD data to external standard codes, enabling the DoD to share interoperable data with the VA, other government agencies and the industry.

Medical Informatics

The DoD is contributing invaluable knowledge regarding real world use of standard terminologies, and is also furthering the development of the standards. In addition, 3M has pioneered tool automation during the standardization project. Processes have been developed to streamline the work and to assure accuracy and consistency.

Ongoing Needs

The DoD has achieved the integration of standard coding schemes and interface terminologies. The challenge remains to incorporate the source lexicons of the commercial software applications used by the DoD. Also, while CHCS is being transitioned into CHCSII, the HDD needs to maintain synchronization with the master files at each MTF for the functions or modules still in operation. This process is called Sustainment Mapping.

The integration, implementation, usage and maintenance of vocabularies present a highly complex long-term effort. Terminologies need to be managed at the enterprise level with proper oversight and decision-making. Automated tools and processes are required for maximum benefit and minimum cost. The Terminology Services Bureau (TSB) will provide an integrated total solution for vocabulary management and maintenance.

Key Features of the Terminology Services Bureau

The Department of Defense Enterprise Reference Terminology as the Core

Standard and source vocabularies need to be integrated with DoD's interface terminologies, forming the DoD Enterprise Reference Terminology (ERT). The foundation has already been provided by the HDD and the DoD standardization accomplishments. The concepts required to encode the data from core administrative and clinical functions are organized in domains (groupings) and linked by relationships to support DoD's application requirements. The ontology of the ERT – the structure and organization of the domains and their interrelationships – is clearly defined and understood. Each new vocabulary to be incorporated must fit within this framework for correctness and applicability, and to achieve economy of scale in effort.

Automated Processes and Tools for Vocabulary Integration

Tool and process automation is another area in which the DoD can capitalize on the accomplishments. While a variety of automated tools have been developed to support mapping as well as vocabulary and knowledge base development, including the management of regular updates from external standards, incorporating source lexicons from software applications poses an added challenge. However, the existing tools can be enhanced with a comparatively small effort to support the ontology definition and instantiation needed for the specialized source lexicons.

Extensibility and Scalability

Extensibility and scalability is the philosophy underlying the design of the HDD table structure and information model. The same principle is emphasized in the TSB approach, so that the maintenance cost will remain stable or even decrease over time. The TSB will apply new technology in the field of Medical Informatics and Computer Science, such as natural language processing or text/voice recognition, to improve the vocabulary processing work. Continuous quality improvement is a critical requirement.

Comparing Existing and Proposed Processes for the Department of Defense

Incorporation of a Software Lexicon

Based on our experience with the PKC prototype project and a subsequent Statement of Work (SOW), the current process appears to divide up the work into incremental projects, resulting in a significant portion of the effort being repeated each time. The proposed process will be to manage the PKC findings as an entire vocabulary, assisted by the automated HDD tools. First, all PKC coupler items will be normalized to a unique set, then grouped into appropriate domain areas for ontological definition. The resultant PKC ontology is compared to the HDD knowledge base for conflict resolution. The PKC ontology is instantiated with the normalized findings, and specific matching tools are applied. Finally, after expert review and quality assurance, the content will be loaded automatically into the HDD.

Sustainment for CHCS Mapping

Changes in the master files, called Deltas, are captured daily for mapping so that the HDD always reflects the current state at the MTFs. The problem is that the deltas at each MTF are managed independently, resulting in a potential duplicate effort of as much as 100 times. The TSB approach will manage all the deltas from all the MTFs in a unified effort for economy of scale. 3M and Integic are developing a shared database to automate the data exchange. At 3M, the delta files from all sites will be integrated into a single Delta Database, replacing the current manual input processes. We are also building the centralized HDD Editing/Mapping Environment to integrate all automated HDD tools and processes. With continued tool enhancement to focus on maintenance, we anticipate decreasing the current sustainment effort to a stable, affordable level.

Roles and Responsibilities

With guidance from the DoD, the TSB will provide the strategic oversight with regards to terminology requirements and prioritization. The TSB has two main responsibilities: vocabulary integration and maintenance. The TSB will focus on the development and utilization of automated tools and processes, and coordinate with Integic and other software and solution providers. The TSB will provide Medical Informatics expertise, participate in standard activities, and liaison to the VA and other government agencies. The ultimate goal of the TSB is to achieve high quality, standardized, interoperable coded data. Towards this end, the TSB will evaluate data and vocabularies, develop strategies and recommendations, and implement solutions.

We propose a Steering Committee with representation from the DoD Program Offices (e.g. CITPO, RITPO, IM), all three branches of military service, and DoD standard committees and initiatives. The DoD may also appoint representatives from the integrator or other software providers, and 3M will provide a Medical Informaticist. The Steering Committee will provide oversight and strategic direction to the Medical Informatics Team, the leadership that directs and coordinate all TSB activities and also provides . Consisting of 3M Medical Informatics experts, a project manager and the HDD team lead, the Medical Informatics Team will also provide representation to DoD, other government agencies, industry, and standard activities.

The 3M HDD department is currently organized into Functional Groups according to domain expertise (e.g. pharmacy, laboratory, tools and processes). These groups will come under the direction of the TSB Medical Informatics Team. As needed, TSB Project Teams will be formed to emphasize particular objectives (e.g. PKC mapping), pulling resources from the appropriate Functional Groups. The use of the existing infrastructure will result in the added advantage of a minimal setup cost for the TSB. Last but not least, representation from other government agencies and industry can be added to the Steering Committee to propagate the TSB as a national body, and to promote the DoD ERT as the national standard for data interoperability.

Benefit Realization

The anticipated return on investment can be estimated from our experience in the CHCSII projects. 3M has focused on tool development and process automation to enhance quality, increase efficiency and reduce cost. From 2000 to present, the DoD contributed an estimated $260,000 to 3M's engineering efforts in tool development. In return, the mapping effort was nearly halved, from 1080 hours per MTF to 720 hours each in 2000 and 560 hours per site in 2003. DoD realized a cost savings estimated at over $3,500,000.

Continuing the focus on quality and efficiency into sustainment, we recently closed a Six Sigma project titled "Minimize HDD Sustainment Mapping Costs for Department of Defense". The project showed that we were at Five Sigma for the proportion of mapping results flagged for questioning during Integic testing (not all mapping errors, but including items not mapped because of outstanding questions or missing information). We are currently automating the data exchange process with Integic, and developing the centralized 3M Delta Database for file processing. We estimate being able to reduce the 3M effort in delta file management from over 1400 hours per month, across all MTFs, to 160 hours per month, a reduction of 89%. This one-time 3M development cost, estimated at approximately $220,000, will result in an anticipated savings of approximately $120,000 per month. We expect the same pattern of economy of scale achieved through tool optimization and process automation for source lexicon incorporation.

Conclusion

The DoD plays a prominent role in the world of the EHR, through its advanced organizational structure, resource abilities, and the reach of its clinical community. The DoD-3M partnership in the TSB builds upon work accomplished and extends the value of DoD's investment in the HDD. The DoD already has a strong foundation in the ERT implemented in the HDD. The automated tools and processes already deployed can be enhanced to focus on vocabulary integration and sustainment. The Medical Informatics expertise and experience as well as knowledge regarding the DoD's system, workflow, data and requirement are already in place. Thus, the TSB approach for vocabulary management is one of maximizing the advantages already gained, and minimizing the integration and maintenance effort to a stable, sustainable level. The goal is to enable the DoD to show the world the power of interoperable data used effectively and efficiently to provide the highest quality care, support population research, and improve cost and outcomes.

THE NEED FOR A TERMINOLOGY SERVICES BUREAU

Health Care Vocabularies

Health care is a large, complex system requiring diverse administrative and clinical functions; hence, many terminologies have been created to support specific purposes or applications. Terminologies can be grouped according to their source, as follows.

1. "Standard Coding Schemes" – Vocabularies developed independent of a software application
   - International Classification of Diseases, 9th Edition, Clinical Modification (ICD9CM)
   - Common Procedural Terminology (CPT)
   - National Drug Code (NDC)
   - Systematized NOmenclature of MEDicine Clinical Terminology (SNOMED CT)
   - Logical Observation Name Identifiers and Codes (LOINC)

2. "Source Vocabularies" – Lexicons embedded within a commercial software application
   - Medcin (from Medicomp)
   - Problem Knowledge Coupler (PKC)
   - National Drug Data Files (NDDF) from First Data Bank (FDB)

3. "Interface Terminologies" – Master files containing enterprise- or facility-specific codes for various Electronic Health Record (EHR) modules
   - Composite Health Care System (CHCS) of the Department of Defense (DoD)
   - Veterans Integrated Services Technology and Architecture (VistA) of the Veterans Administration (VA)

Data Interoperability and Standardization

The above examples are used by, or of interest to, the DoD, together with many other vocabulary sets. Encoding the patient data with disparate, non-communicating vocabularies has resulted in data that is not interoperable among applications, across facilities, or over time. For instance, CHCS has been in use throughout all DoD Military Treatment Facilities (MTFs) for nearly a decade. Each MTF has a set of master files used to encode patient data in CHCS. The content of each master file is not identical across facilities (e.g. "1" may mean "Acetaminophen" at an MTF but "Aspirin" at another). The data is thus not encoded the same way from one DoD MTF to another, and not interoperable.

Interoperability means that the data encoded at one site is interpretable at another site, as if the data were encoded there. Interoperability allows data to be used in applications regardless of origin, and to be aggregated and compared across location and time. One means to achieve interoperability is data standardization. Data standardization refers to the use of the same set of codes to encode data throughout a system. As an example, for the domain of "sex", one may decide always to code the sex of male as "1", female as "2", and unknown as "3". The domain of "sex", consisting of three members, "male", "female" and "unknown", forms a vocabulary, albeit a very simple one. If all data about sex is coded consistently according to this vocabulary, the data should always be understandable and usable for analysis across time and location.

Department of Defense Accomplishments

The DoD has made a commitment to data interoperability through standardization. Four years ago, the DoD began the transition to the next-generation system, CHCSII. The DoD selected the 3M Healthcare Data Dictionary (HDD) to accomplish data interoperability in CHCSII and with the health care industry at large. The content of the HDD is built with industry-standard vocabularies. Concepts in the HDD are each identified by a Numerical Concept IDentifier (NCID). NCIDs are used for enterprise-wide encoding of clinical data for storage in the CHCSII Clinical Data Repository (CDR). Standard vocabularies are incorporated in the HDD. External standard codes, such as LOINC, are mapped to NCIDs. Through mapping, the HDD can translate between one standard and another, between legacy systems, and between a legacy system and a standard. To exchange data with external systems, such as the VA, the HDD can translate the DoD data from NCIDs to the requested external standard code (e.g. LOINC).

The CHCS master files are mapped to the HDD. When content is loaded into a CHCS master file (e.g. a list of insurance companies), a unique Internal Entry Number (IEN) is automatically generated as the item identifier (primary key). The identical item from the same file will receive different IENs in different MTFs, unless the same IEN is assigned by a rare coincidence. For interoperability, the different IENs for the same item need to be mapped to a single standardizing concept. To date, a total of 3 million CHCS items from the five domain areas of Demographics/Encounters, Laboratory, Microbiology, Pharmacy and Radiology/Text Reports have been mapped for all 101 MTFs. Because NCIDs instead of IENs are used to encode CHCS data in the CHCSII CDR, previously isolated islands of legacy data can now be interoperable across all MTFs, and across time.

The CHCSII data standardization accomplished by the DoD is unprecedented in scale and implication, and DoD should be applauded for its commitment. The DoD will be able to share correctly and consistently coded data for patient care among MTFs, greatly contributing to care delivery and quality. The coded data will also enable accurate, efficient and useful decision support and patient safety efforts. Because the HDD is able to translate from NCIDs to whichever standard codes are requested (e.g. LOINC), the DoD is in a leading position for data interoperability in communicating with the rest of the world. For the benefit of DoD's health care beneficiaries and veterans, the DoD will be able to share interoperable data with the VA. The DoD has laid its foundation for population based applications that require coded, enterprise-wide data, e.g. data warehousing and outcomes research. If the standardization work has not been done before the data enters the warehouse or is used in population studies, a great deal of cost and effort would be required to manipulate the data, and would yield results of variable quality.

The DoD is contributing invaluable knowledge regarding real world use of standard terminologies, not to mention contributing to the further development of the standards themselves. For instance, the size of the laboratory file from the DoD MTFs ranged from about 1,000 to over 18,000 rows of data, averaging about 5,000 rows. These approximately 500,000 rows from all 101 MTFs have been mapped to just fewer than 20,000 NCIDs. Over half of these DoD laboratory results do not currently have a LOINC code. 3M is preparing them for submission to LOINC. The latest release of LOINC in October 2003 contains approximately 20,000 laboratory codes and roughly 14,000 clinical observation codes. If only half of the submission receives a LOINC code, then the DoD would have grown the size of the LOINC laboratory database by 25%, a significant enrichment.

The DoD has also contributed to the field of Medical Informatics vocabulary research through the tool and process automation pioneered by 3M during the standardization project. Over the past few years, a variety of automated tools have been developed to support mapping as well as vocabulary and knowledge base development. The lessons learned have been shared with the industry through presentations and publications. These technical advances made the data standardization feasible, and also made possible the real world use of standard vocabularies. The tools and techniques include:

- Matching Assistant – general mapping tool for any candidate domain, calling upon the HDD for source domains. Utilizes lexical techniques such as di-gram matching as well as parsing, synonymy, text/keyword searching, and other Natural Language Processing methods.

- Vocabulary Browser – reviews concepts, representations, contexts and relationships (starting point is concept); performs adds and updates.

- Relationship Editor – analyzes HDD relationships; performs adds and deletes. Starting point is domain.

- Domain specific mapping tools for the most labor-intensive areas – laboratory/microbiology, pharmacy, insurance and location. Performs automated mapping to existing HDD content; creates new concepts, representations, contexts and relationships for addition into the HDD. Includes functionality for content maintenance (e.g. monthly pharmacy updates from FDB). In addition to lexical techniques, uses innovative semantic matching techniques developed by 3M. For example, the Lab Editor incorporates semantic transformation rules, attribute comparison, and decision factors (see reference).

- HDD Loader – performs automated creation of concepts, representations, contexts and relationships in the HDD; called upon by the mapping tools.

Last but not least, processes have been developed to streamline the standardization work for efficiency, and to assure accuracy and consistency. Expert review and quality assurance are part of the mapping procedure. Factors that would affect a mapping judgment are defined and documented in operating procedures. Examples of the factors are granularity, composition, and specificity attributes (e.g. place or person name associated with an item). The automated mapping tools also incorporate the factors and rules. In addition, a self-learning function is included in the mapping tools. An expert mapper will review those items that the tool was not able to map automatically. The mapping decisions and results are fed into the self-learning function. The tool thus improves its automated mapping with every run. A Mapping Quality Assurance database further ensures mapping accuracy by providing a detailed history log of all the questions asked, comments made and decisions taken.

Ongoing Needs

Through the tasking to 3M, the DoD has achieved the integration of the first and third groups of terminologies described earlier: "Standard Coding Schemes" and "Interface Terminologies". Two challenges remain. First, the DoD uses multiple commercial software applications, each with its own "Source Vocabulary" or lexicon embedded within. Examples are Medcin (from Medicomp) and Problem Knowledge Coupler (PKC). This group of terminologies also needs to be integrated for complete data interoperability. Second, while CHCS is being transitioned into CHCSII, the HDD needs to maintain synchronization with the master files at each MTF for the functions or modules still in operation. The process, called Sustainment Mapping, involves capturing the changes in the content of the master file and making the corresponding, appropriate changes in the HDD. The goal is to have the HDD reflect the current state at the MTFs, so that data are encoded correctly in the CDR.

The importance of having standardized, interoperable data cannot be overemphasized. Vocabularies play a critical role in deciding the quality of encoded data. Integrating, implementing and using vocabularies represent a highly complex effort. Together with the necessity for long term maintenance, these interwoven factors drive the need to manage terminologies at an overarching, enterprise level. Proper oversight and decision making, taking into account the needs of the entire organization, is required to serve the DoD over the entire life cycle of CHCS/CHCSII. Automated tools and processes are required to integrate and maintain the vocabularies for maximum benefit and minimum cost. The vocabulary management and maintenance, as a well planned, coordinated and efficient whole, will be a total solution provided by the Terminology Services Bureau (TSB).

THE TERMINOLOGY SERVICES BUREAU PROCESS

Key Features of the Terminology Services Bureau

The Department of Defense Enterprise Reference Terminology as the Core

Standard and source vocabularies need to be integrated with DoD's interface terminologies, forming the DoD Enterprise Reference Terminology (ERT), to ensure comprehensiveness and applicability to all the data needs across MTFs. The foundation is already provided for the DoD by the HDD, through data standardization for CHCS. The concepts required to encode the data from core administrative and clinical functions are organized in domains (groupings) and linked by relationships to support DoD's application requirements. The ontology of the ERT – the structure and organization of the domains and their interrelationships – is clearly defined and understood. Each new vocabulary to be incorporated must fit within this framework. Properly integrated, the candidate vocabulary can add refinements and enhancements to the ERT ontology without damaging its basic structure or consistency. Conflicts must be resolved without changing the ERT ontology to the extent that it is no longer compatible with current functions or previously encoded data. This is why we stress the need for the DoD ERT as the foundation. An ERT built from all the data elements used by the enterprise, and supporting all the functions of the enterprise, will provide the template upon which to add new vocabularies. Using the ERT as the core for vocabulary integration creates a stable reference. Without it, the addition of each new vocabulary will cause a drift in content, and there is also no economy of scale; each new "merge" will result in a similar effort to the previous. In contrast, vocabulary integration into an ERT takes advantage of the previous work done, resulting in a much more cost-effective approach.

Automated Processes and Tools for Vocabulary Integration

The work to incorporate a new vocabulary and to maintain it long term is non-trivial. Obviously, it would be inefficient and costly to attempt it with manual effort only. Automated tools and processes are required to capitalize on the vocabulary and knowledge base already provided by the HDD. Automation has the additional advantage of imposing consistency and integrity on the vocabulary content. Human expertise should be used where it is most critical and where creativity or complex reasoning is required – in the design and development of the automated tools and processes, and in the review of their output to provide continuous improvement.

Tool and process automation is another area in which the DoD can capitalize on the work already done by 3M. Over the past few years, a variety of automated tools, as described earlier, have been developed to support mapping as well as vocabulary and knowledge base development. They are also used to incorporate regular updates from the external standards – the release is compared to the HDD, new content is mapped automatically, reviewed by experts in the quality assurance process, and loaded (also automatically). Lessons learned are used for continuing improvement and enhanced automation. We are in the process of unifying and further automating the HDD tools into a single editing environment for ongoing HDD development and maintenance.

Incorporating source lexicons from software applications poses an added challenge. The lexicons tend to be highly specialized and most would not follow vocabulary principles (see reference). For instance, our experience with the NDDF database from FDB led us to develop a customized program to manipulate and load its pharmacy lexicon into the HDD. Run monthly (or as often as required), the program also provides an update function that compares new NDDF data to the HDD. It loads the new content and applies matching logic to the changes. Those changes that cannot be resolved algorithmically (e.g. NDC reuse – same NDC, different drugs) is presented for human review in an easy to use Graphical User Interface. The review result is loaded automatically (new concepts created, new representations added, invalid concepts set to obsolete and outdated representations “inactivated”).

Applying the lessons learned and capitalizing on the existing tool capabilities, the TSB will expand the Relationship Editor to include an Ontology Editor module. The Ontology Editor will support customized incorporation of source lexicons from software applications. It will provide for the definition of domains and classes, and their interrelationships. Inference will be used to streamline the ontology construction (e.g. if Roses are Flowers and Flowers are Plants then Roses are Plants) and ensure correctness (e.g. no circular relationships – A is parent of B and B is parent of A).

The next step is to instantiate the derived ontology with all the concepts from the vocabulary, making the necessary adjustments. Instantiation – populating the ontology (abstract definitions) with actual concepts – is a critical requirement if the vocabulary is to be used in operation and not to remain merely theoretical. Instantiation is performed automatically, again streamlined with the use of inference, and the tool will present conflicts between the vocabulary content and the ontology. Adjustments are made until the ontology and the instantiated concept relationships are a consistent whole.

Once the ontology instantiation is completed, other automated tools are called upon to compare the concepts and relationships of the new vocabulary to that already in the HDD. Again, conflicts between the new vocabulary and the HDD are presented for adjustment and resolution. Automated tools perform the mapping to existing HDD concepts or create new concepts and relationships, and load the content from the new vocabulary into the HDD.

Extensibility and Scalability

The TSB approach must be extensible and scalable, so that the maintenance cost will remain stable or even decrease over time. Ongoing quality review must be part of the vocabulary management, to maintain the integrity of the patient data, and to avoid the cost of fixing errors. A key responsibility of the TSB will be to maintain the accuracy and consistency of the ERT in a cost-effective manner. The TSB will apply new technology in the field of Medical Informatics and Computer Science, such as natural language processing or text/voice recognition, to improve the vocabulary processing work. Continuous quality improvement is a critical requirement.

Extensibility and scalability is the philosophy underlying the design of the HDD table structure and information model. The HDD tables are designed to provide maximum flexibility for the content of the vocabulary and knowledge base. The table structures and the information models do not constrain the content. The flexibility allows the complexity of medical knowledge to be represented. 3M will bring the experience and knowledge gained from the HDD into the TSB.

Comparing Existing and Proposed Processes for the Department of Defense

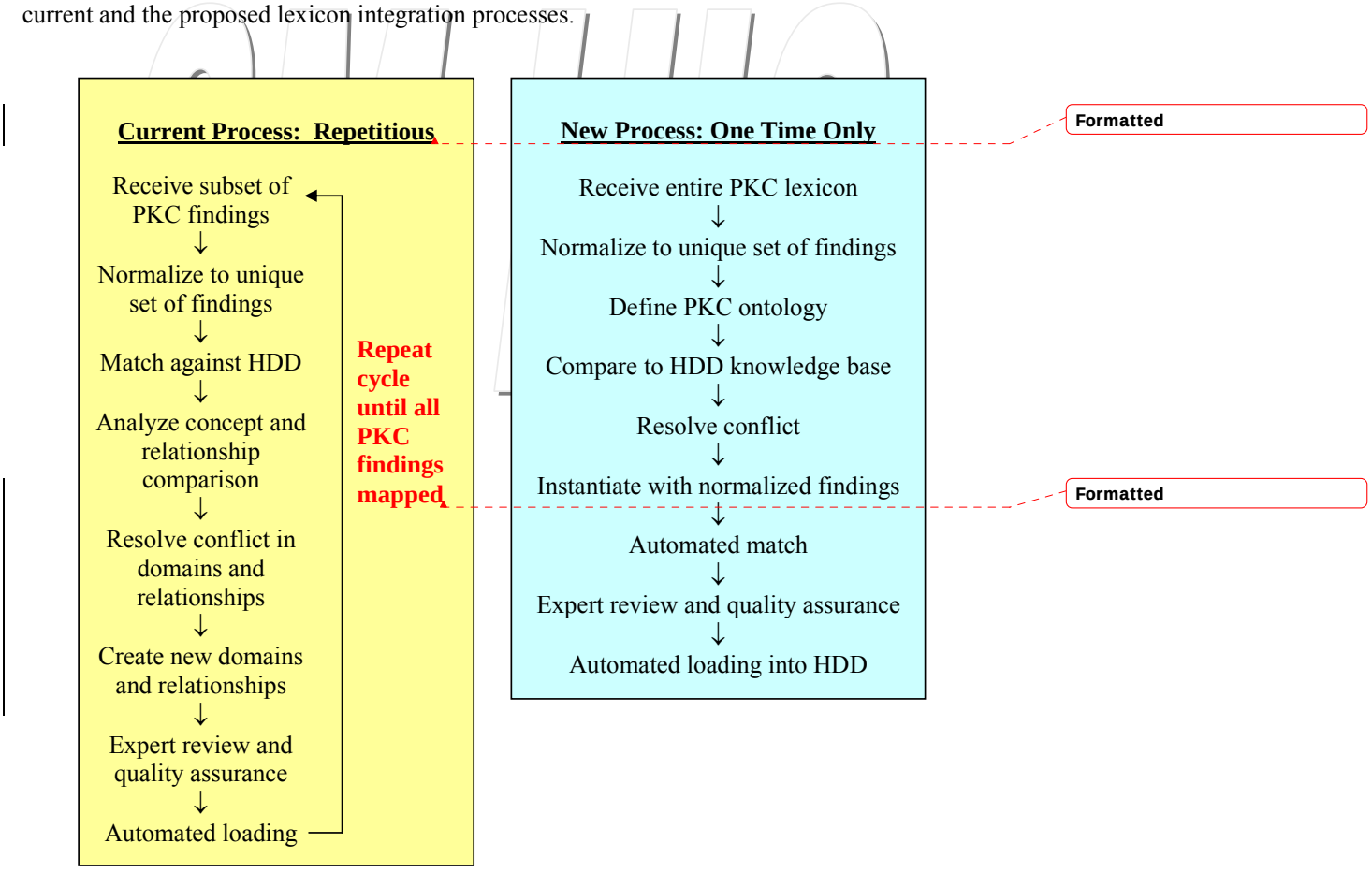
Incorporation of a Software Lexicon

Towards the end of 2002, 3M took part in a prototype project to integrate the Diabetes Mellitus Management and the Health Evaluation Assessment Review (HEAR) couplers from the Problem PKC application into CHCSII Encounter workflow. It was found that the "theoretical" mapping of PKC laboratory results to the standard vocabulary of LOINC could not ensure the correct retrieval of all site data from the MTFs. Therefore, 3M was given a small list of 10 items from the coupler and asked to demonstrate the correct way to map them into the HDD. The results are shown in the following table. The PKC coupler questions are included in Appendix A. The experience illustrated the necessity of the DoD ERT as the foundation for terminology integration.

PKC Coupler	Question #	Item for Mapping	3M HDD NCID	Explanation
Diabetes	Q81	Beta Blocker	3000250309	NCID is for the Drug Class. Data stored in the CDR may be Drug Class, Ingredient, Generic Medication or Drug Product, depending the granularity of the data that has been sent to the CDR. Ingredient, Generic Medication and Drug Product concepts will have a HasMember relationship to this Drug Class and can be retrieved with this NCID.
Diabetes	Q4	Metformin	3000251090	
Diabetes	Q4	Insulin	3000250167	
Diabetes	Q4	Sulfonylurea: 2nd Generation	Does not currently exist	FDB does not divide sulfonylureas by generations. The HDD can create the subdomains with the appropriate content and provide the NCID for 2nd generation sulfonylurea.
HEAR	Q49/50	Laxatives	3000250231	Drug Class of Laxatives and Cathartics. Note that both NCIDs should be used to retrieve domain members. A new NCID can be provided to group all laxatives, with these two NCIDs as subdomains.
			3000250397	Drug Class of Local/Rectal Laxatives.
Diabetes	Q6	Allergy to Insulin	3000250167	Note that although this is the same NCID for insulin, the information model in the CDR from which this data should be retrieved is different from that for insulin as the medication.
Diabetes	Q6	Insulin Resistance due to Insulin Antibodies	To be created	There are multiple lab results that can be mapped. A subdomain will be created to group them.
HEAR	Q120/121	Pap Smear	To be created	There are multiple lab results that can be mapped. A subdomain will be created to group them.
Diabetes	Q38	Albumin/ Creatinine Ratio (mcg/mg): Spot Collection	20912	This is a lab result defined by LOINC (code 9318-7) and should be retrieved from the appropriate element in the lab information model.
HEAR	Q134	Mammogram	84550	This is a text report subdomain created to group the multiple mammogram reports. It is not the test order.

Recently, 3M responded to a Statement of Work (SOW) to map an additional 150 PKC items from the Diabetes Mellitus Management Coupler into the HDD, in the categories of laboratory, medication, allergy, and radiology results. Assuming all PKC findings will be mapped eventually, this process as it appears to us does not maximize efficiency or minimize cost. It is logical and prudent to do a proof of concept to gain understanding and experience, as in the prototype project. However, subsequent incremental projects that divide the work up will result in repeating a significant portion of the effort each time. For example, a set of medications will be mapped each time, and new subdomains created. It is very likely that the same medication finding will be repeated in multiple couplers and thus sent for mapping multiple times.

The proposed process will manage the PKC findings as an entire vocabulary, assisted by the automated HDD tools. First, all PKC coupler items will be analyzed together and distilled down to unique items for matching. They are then grouped into appropriate domain areas for ontological definition. Once the classes and relationships are defined, the resultant PKC ontology is compared to the HDD knowledge base. Any conflict with the HDD semantic net is resolved, and new domains and relationships to be created in the HDD are highlighted. Then, the PKC ontology is instantiated with the normalized findings, and specific matching tools are applied according to the domain area and fit. Finally, after expert review and quality assurance, the content will be loaded automatically into the HDD. The following figures compare the current and the proposed lexicon integration processes.



Once loaded into the HDD, PKC updates will also be maintained regularly. The HDD toolset and mapping environment are already set up to manage updates from external vocabularies. The latest release of the lexicon can be compared to the state of the HDD, and the changes presented for review and subsequent automated loading.

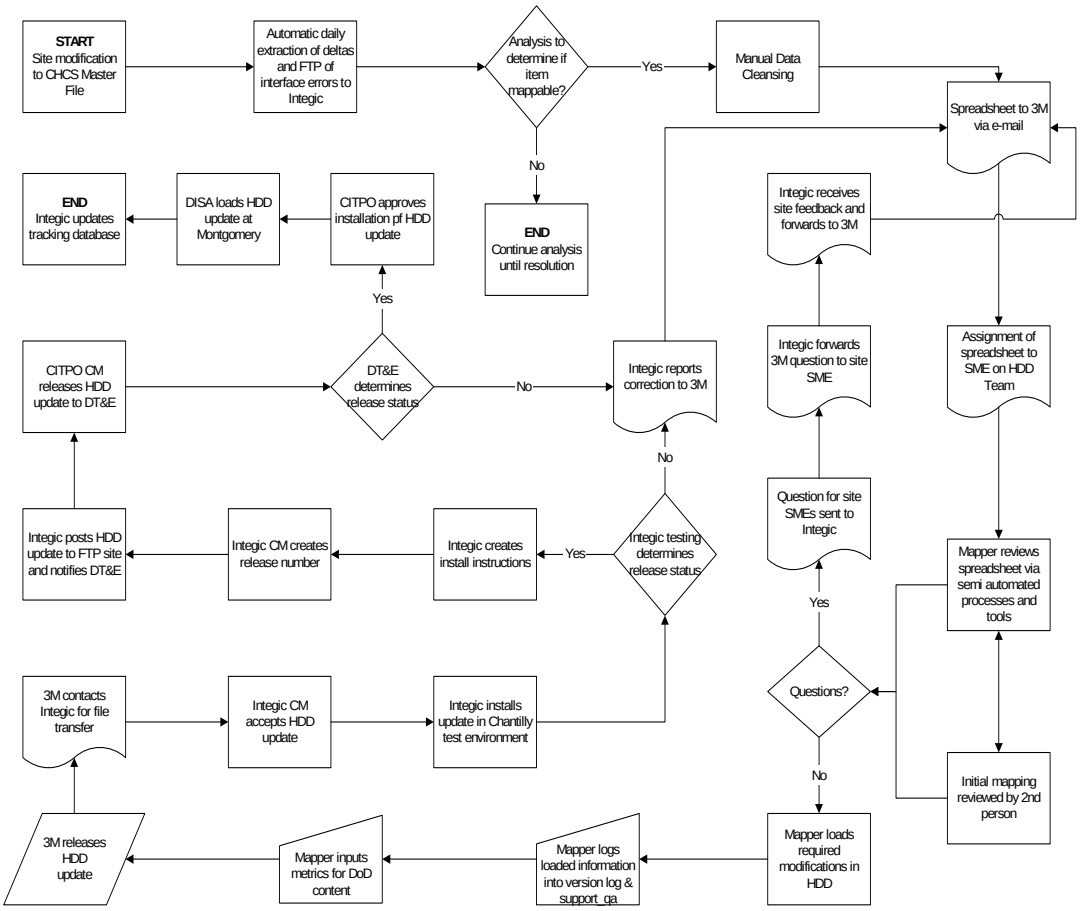
Sustainment for CHCS Mapping

Now that the CHCS master files have been mapped to the 3M HDD, to maintain synchronization, changes in the master files, called Deltas, are captured daily for review. Corresponding updates are made in the HDD so that the DoD Reference Terminology always reflects the current state at the MTFs. New interface codes, usually the CHCS Internal Entry Number (IEN), in a delta file are mapped to appropriate concepts in the HDD. If an IEN already exists, then the delta is reviewed to see if the meaning of the entry has changed. For example, IEN 123 used to mean "Acetaminophen", but the delta now shows it as "Aspirin". In the HDD, IEN 123 must be dissociated from the concept of Acetaminophen, and instead mapped to the concept of Aspirin. Then, when the IEN comes across as the interface code in a transaction from that particular MTF, the correct concept will be stored in the CDR.

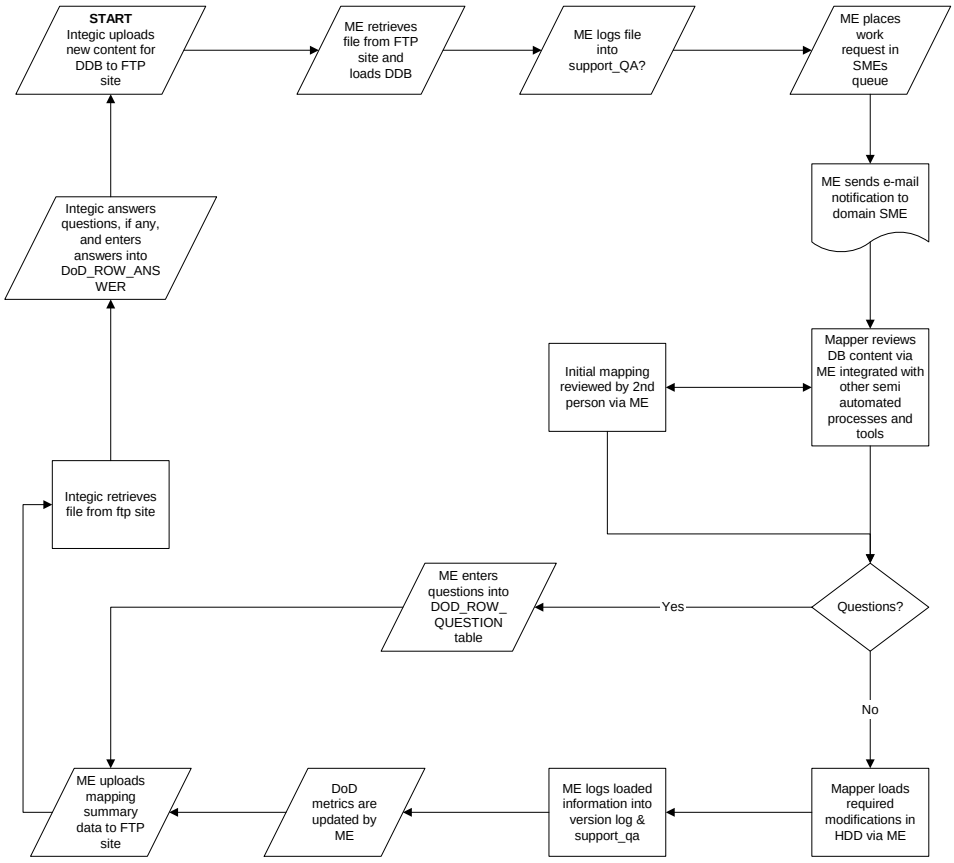
Terminologies and code sets do evolve and need to be maintained to meet clinical requirements. However, for the DoD, the problem is that the deltas at each MTF are managed independently. Across all MTFs, the mapping work is potentially duplicated as much as 100 times, driving sustainment cost up. The TSB will manage all the deltas from all the MTFs in a unified effort for economy of scale. 3M and Integic are developing a shared database to automate the data exchange. At 3M, the delta files from all sites will be integrated into a single Delta Database for processing. Current manual input processes will be replaced. We are also building the centralized HDD Editing/Mapping Environment as the management tool to integrate all automated HDD tools and processes. With continued tool enhancement to focus on maintenance, we anticipate being able to decrease the current effort for the DoD sustainment to a stable, affordable level.

Automating the data exchange and the file management, so that the delta requests from all sites on a single day can be treated as one, is an immediate solution that is already in progress. The current manual process for sustainment is shown in the first figure on the following pages. The portion being automated and improved is shown in the second figure. In addition, as the Terminology Services Bureau, we are prepared to propose innovations that would advance the state of DoD data standardization. An example for sustainment is given in Appendix B. We understand that business and logistic factors may not allow the DoD to adopt all technical ideas, but we expect to use our Medical Informatics expertise to explore solutions that may be advantageous to the DoD.

Current Sustainment Process – From Site Delta to HDD Update Application:



Improvement to the Current Sustainment Process: Automating File Management with the 3M Delta Database (DDB) and Mapping Environment (ME)



ROLES AND RESPONSIBILITIES OF THE TERMINOLOGY SERVICES BUREAU

We envision the TSB as a team of experts in data quality and vocabulary implementation working with the DoD, Integic as the integrator, and other software and solution providers. With guidance from the DoD, the TSB will provide the strategic oversight with regards to terminology requirements, enabling the prioritization of vocabulary projects to meet DoD's needs. Both the enterprise perspective as well as MTF level input will be solicited. Together with the DoD, the TSB will also set cost and timeline objectives. To maintain currency with DoD's operation and requirements, and to support future growth and evolution, the TSB should participate in DoD's standardization committees, initiatives and other activities. Where applicable, the DoD can provide subject matter expertise. An example is the dental information modeling project that also created the dental lexicon in the HDD. All specialty areas of military dentistry with equal representation from all three service branches cooperated in the development and evaluation.

The TSB has two main responsibilities: vocabulary integration and maintenance. To accomplish the work with maximum efficiency and quality and minimum cost, the TSB will focus on the development and utilization of automated tools and processes. The TSB will work with Integic and other software and solution providers to provide vocabulary coordination and support to the applications. The TSB will provide Medical Informatics expertise and support to the DoD. On behalf of the DoD, the TSB will participate actively in standard activities in the health care industry. The TSB will provide liaison to the VA and other government agencies in the pursuit of interoperability in data exchange.

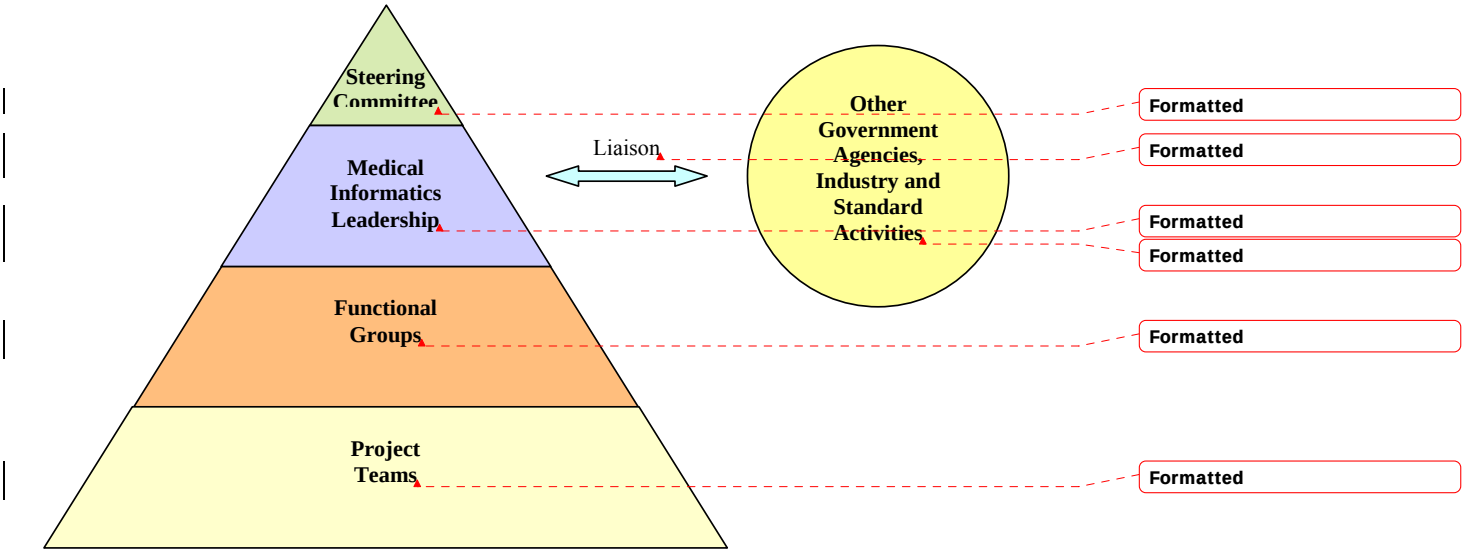
The ultimate goal of the TSB is high quality, standardized, interoperable coded data for the DoD. Towards this end, the TSB will evaluate data and vocabularies, develop strategies and recommendations, and implement solutions. For instance, dental data is not stored in the CDR currently; the dental information model and lexicon are not used. The dental information is thus not integrated with the medical data. The annual dental encounter requirement for all service personnel is a valuable source for general medical information; it provides significant opportunities to collect information about medical condition, to deliver patient education, and to encourage primary prevention of disease. The TSB will review the current dental storage schema, the master file codes, and a sample of data being stored. The findings will be evaluated against the dental information model and lexicon. A proposal can then be put before the DoD to include recommendations, scope of effort, and risk and benefit analysis, as well as the coordination and effort needed for software reengineering and performance assessments.

The following figure illustrates the suggested organizational structure of the TSB. We propose a Steering Committee with representation from the DoD Program Offices (e.g. CITPO, RITPO, IM) and all three branches of service. Participation from standard committees and initiatives is strongly encouraged. The DoD may also appoint representatives from the integrator or other software providers, and 3M will provide a Medical Informaticist to the committee. Regular, monthly meetings are expected to maintain oversight and clarity regarding strategic intent.

The next organizational layer is the Medical Informatics Team. This is the leadership that directs and coordinates all TSB activities under the guidance of the Steering Committee. It will consist of 3M Medical Informatics experts, a project manager and the HDD team lead. The Medical Informatics Team is also the representation to DoD, other government agencies, industry, and standard groups (e.g. Health Level 7 – HL7).

The 3M HDD department is currently organized into Functional Groups according to domain expertise (e.g. pharmacy, laboratory, tools and processes). These groups will come under the direction of the TSB Medical Informatics Team. As needed, TSB Project Teams will be formed to emphasize particular objectives (e.g. PKC mapping), pulling resources from the appropriate Functional Groups. Matrix management with regular meetings will ensure coordination. The HDD group infrastructure has been fine tuned and found to be efficient and effective. The experienced Medical Informatics, project management, mapping and vocabulary implementation talents are already in place. This is an added advantage to the DoD, in that there would be no additional setup cost in utilizing the existing infrastructure.

A final point to consider: the TSB is laying the groundwork to propagate the DoD ERT as the national standard for data interoperability. The organizational structure described above will accommodate the growth of the TSB into a national body. Representation from other government agencies and industry can be added to the Steering Committee to provide liaison and national oversight. The DoD is thus poised to lead the country to manage terminologies in data standardization and interoperability.



BENEFIT REALIZATION

The clinical benefits of having high quality, coded, standardized and interoperable data are well understood. In this section, we will estimate the return on investment from an economic perspective. We will base the discussion on our experience in the CHCSII projects.

Throughout the CHCSII mapping projects for the DoD MTFs, 3M has focused on tool development and process automation to enhance quality, increase efficiency and reduce cost. At the beginning of 2000, it was estimated that automated tools would reduce the effort estimate for mapping from 1080 hours per MTF to 720 hours. The reduction has continued over time as the tools continued to be improved and automation enhanced, to an estimated 560 hours per site in 2003. Altogether, from 2000 to present, the DoD contributed an estimated $260,000 to 3M's engineering efforts in tool development. In return, the mapping effort was nearly halved. DoD realized a cost savings estimated at over $3,500,000 (after spending on tools). In addition, timeline compression has allowed completion of all initial MTF mapping in 2003, preparing the sites to enter sustainment.

3M has maintained a high level of quality and maximize efficiency while minimizing cost. This focus is continuing into mapping sustainment. In the fall of 2003, we closed a Six Sigma project titled "Minimize HDD Sustainment Mapping Costs for Department of Defense". With Integic participation, the findings and recommendations were presented to the DoD. The project showed that we were at Five Sigma for the proportion of mapping results flagged for questioning during Integic testing. These were not necessarily all mapping errors, but include items not mapped because of outstanding questions or missing information, disagreement in interpretation, etc. We are currently progressing on the improvement efforts developed in the project. Two of them are the automation of the data exchange process for sustainment mapping with Integic, and the centralized 3M Delta Database for file processing, as mentioned earlier. Testing will begin in March 2004. Once these are in place, we estimate being able to reduce the 3M effort in delta file management from over 1400 hours per month, across all DoD MTFs, to 160 hours per month, a reduction of 89%. The 3M development cost is estimated to be just over $220,000. This one time investment will result in an anticipated savings of approximately $120,000 per month.

We expect to show the same pattern of economy of scale achieved through tool optimization and process automation for source lexicon incorporation. For instance, in the PKC prototype project, the 3M work effort was proportionately small (estimated at 30 hours at the start of the project). Similarly, the follow on SOW provided for 150 hours of mapping. However, these labor estimates should not be used to project the effort to incorporate all PKC couplers. Simply multiplying the prototype effort by the number of couplers would result in an incorrectly large labor amount. The manual review, especially given the small list of items, is to demonstrate the feasibility of incorporating PKC couplers into the HDD. The actual approach to incorporate the full lexicon into the HDD would be as outlined earlier for the Terminology Services Bureau. All PKC coupler items would be analyzed together and grouped into appropriate domain areas for ontological definition. Reviewing the candidate vocabulary as a whole, with the help of automated tools, will achieve the economy of scale that is not called for during the prototype demonstration.

CONCLUSION

The DoD plays a prominent role in the world of the EHR, through its advanced organizational structure, resource abilities, and the reach of its clinical community. The DoD-3M partnership in the TSB builds upon work accomplished and extends the value of DoD's investment in the HDD. The DoD already has a strong foundation in the ERT implemented in the HDD. The automated tools and processes already deployed can be enhanced to focus on vocabulary integration and sustainment. The Medical Informatics expertise and experience as well as knowledge regarding the DoD's system, workflow, data and requirement are already in place. Thus, the TSB approach for vocabulary management is one of maximizing the advantages already gained, and minimizing the maintenance work and cost to a stable, sustainable level. The goal is to enable the DoD to show the world the power of interoperable data used effectively and efficiently to provide the highest quality care, support population research, and improve cost and outcomes.

3M HDD REFERENCES

Lee Min Lau, Kate Johnson, Kent Monson, Siew Hong Lam, Stanley M. Huff. A Method for the Automated Mapping of Laboratory Results to LOINC. Journal of the American Medical Informatics Association, Proceedings of the 2000 Annual Symposium.

Lee Min Lau, Siew Hong Lam. Applying the Desiderata for Controlled Medical Vocabularies to Drug Information Databases. Journal of the American Medical Informatics Association, Proceedings of the 1999 Annual Symposium.

Lee Min Lau, Stanley M Huff, Roberto A Rocha. An Experience Using LOINC in a Clinical Application. Journal of the American Medical Informatics Association, Proceedings of the 1996 Annual Symposium.

Lee Min Lau, Siew Hong Lam, Stanley M. Huff. Comparing Encounter and Demographics Data Elements among Different Healthcare Enterprises Using a Common Data Dictionary. Journal of the American Medical Informatics Association, Proceedings of the 1998 Annual Symposium.

Lee Min Lau, Siew Hong Lam. Comparing the Mapping of Laboratory Results to LOINC among Different Healthcare Institutions Using a Common Data Dictionary. Journal of the American Medical Informatics Association, Proceedings of the 1999 Annual Symposium.

Patricia S Wilson, Lee Min Lau, Stanley M Huff. Issues Encountered While Standardizing Diagnostic Imaging Procedures of Multiple Healthcare Institutions To LOINC. Journal of the American Medical Informatics Association, Proceedings of the 2000 Annual Symposium.

Shaun C Shakib, Elva Knight, Lee Min Lau. Challenges of Mapping Microbiology Data to a Common Data Dictionary. Journal of the American Medical Informatics Association, Proceedings of the 2000 Annual Symposium.

Edward M Cassin, Lee Min Lau. Strategic and Management Lessons Learned in the Implementation of A Data Dictionary. Journal of the American Medical Informatics Association, Proceedings of the 2000 Annual Symposium.

Joni Endo, Elva Knight, Patricia Wilson, Lee Min Lau. Providing Data Dictionary Support to Multiple Healthcare Enterprises. Journal of the American Medical Informatics Association, Proceedings of the 2001 Annual Symposium.

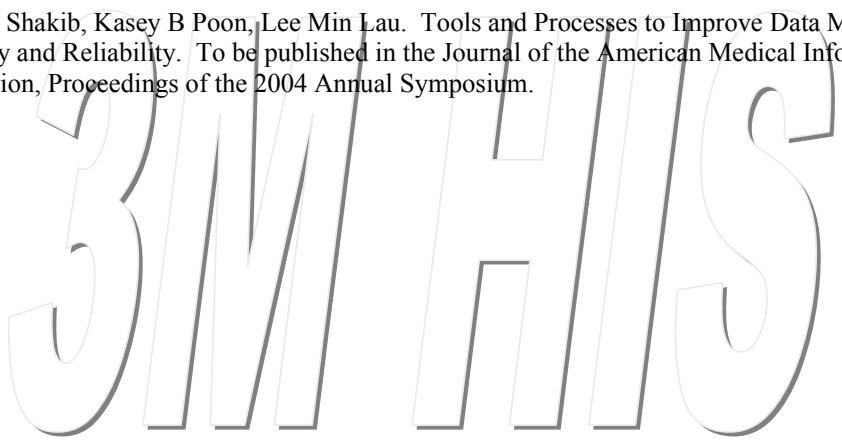
Lee Min Lau, Siew Hong Lam, Steven Barlow, Chuck Lyon, Dale Sanders. Enhancing an Enterprise Data Warehouse with a Data Dictionary. Journal of the American Medical Informatics Association, Proceedings of the 2001 Annual Symposium.

Shaun C Shakib, Joni Endo, Lee Min Lau, J William Karitis. Development of an Integrated Dentistry Information Model. Journal of the American Medical Informatics Association, Proceedings of the 2001 Annual Symposium.

Shaun Shakib, Elva Knight, Joni Endo, Lee Min Lau. An Application to Integrate the Logistical and Technical Aspects of Data Dictionary Support to Multiple Healthcare Systems. Journal of the American Medical Informatics Association, Proceedings of the 2002 Annual Symposium.

Kasey B Poon, Shaun C Shakib, Lee Min Lau. A Quality Assurance Browser Tool for Tracking Vocabulary Mapping. To be published in the Journal of the American Medical Informatics Association, Proceedings of the 2004 Annual Symposium.

Shaun C Shakib, Kasey B Poon, Lee Min Lau. Tools and Processes to Improve Data Mapping Accuracy and Reliability. To be published in the Journal of the American Medical Informatics Association, Proceedings of the 2004 Annual Symposium.



Appendix A. Questions provided to 3M for the prototype project to integrate the Diabetes Mellitus Management and the Health Evaluation Assessment Review (HEAR) couplers from the Problem Knowledge Coupler (PKC) application.

Q4 — Diabetes medications currently taken

- ☐ sulfonylurea: 1st generation
- ☐ sulfonylurea: 2nd generation
- ☐ acarbose (Precose)
- ☐ metformin (Glucophage)
- ☐ miglitol (Glyset)
- ☐ glitazone
- ☐ insulin
- ☐ repaglinide (Prandin)
- ☐ nateglinide (Starlix)

Q6 — Insulin use history

- ☐ allergy to insulin (including human)
- ☐ insulin resistance due to insulin antibodies
- ☐ incapacitating insulin-related problems, failed therapy

Q81 — Current medications

- ☐ ACE inhibitor
- ☐ aminoglycoside antibiotic
- ☐ anabolic steroid
- ☐ antacid containing aluminum
- ☐ anticholinergic drug
- ☐ anticoagulant
- ☐ antihistamine H2 blocker
- ☐ appetite suppressant
- ☐ aspirin or other salicylate
- ☐ azole antifungal
- ☐ barbiturate
- ☐ benzodiazepine
- ☐ beta blocker
- ☐ bile acid sequestrant
- ☐ bronchodilator, adrenergic
- ☐ calcium channel blocker
- ☐ corticosteroid

- ☐ digestive enzyme preparation
- ☐ diuretic therapy
- ☐ erythromycin
- ☐ fluoroquinolone antibiotic
- ☐ hepatic enzyme inducer
- ☐ hepatic enzyme inhibitor
- ☐ hydantoin anticonvulsant
- ☐ MAO inhibitor
- ☐ nonsteroidal anti-inflammatory drug
- ☐ selective serotonin reuptake inhibitor (SSRI)
- ☐ statin (HMG-CoA reductase inhibitor)
- ☐ sulfonamide (sulfa drug)
- ☐ sympathomimetic drug
- ☐ thyroid hormone
- ☐ tricyclic antidepressant

Q38 — Urine

- ☐ proteinuria
- ☐ ketonuria Enter any ONE of the following:

_____ albumin/creatinine ratio (mcg/mg): spot collection:

_____ albumin excretion rate (mg/24 hours): 24-hour collection:

_____ albumin excretion rate (mcg/minute): timed collection:

Q134 - A mammogram is an x-ray of each breast to look for breast cancer. How long has it been since your last mammogram?

- ☐ Less than 1 year ago
- ☐ 1 year but less than 2 years ago
- ☐ 2 years but less than 3 years ago
- ☐ 3 years but less than 5 years ago
- ☐ 5 or more years ago
- ☐ Never
- ☐ Don't know/Not sure

Q49/50 -Do you take any other over-the counter, non-prescription medications regularly (3 times per week or more)? (check all that apply)

- ☐ Antihistamines
- ☐ Decongestants (oral)
- ☐ Decongestants (nasal sprays such as Afrin)
- ☐ Sleep aids
- ☐ Laxatives
- ☐ Aspirin (unless instructed by a health care provider)
- ☐ Non-aspirin pain relievers

Q120/121 - A Pap smear is a test for cancer of the cervix. How long has it been since you had your last Pap smear?

- ☐ Less than 1 year ago
- ☐ 1 year but less than 2 years ago
- ☐ 2 years but less than 3 years ago
- ☐ 3 years but less than 5 years ago
- ☐ 5 years or more
- ☐ Never
- ☐ N/A I don't have a cervix
- ☐ Don't know/Not sure

Appendix B. Innovation to Improve Sustainment Management

The TSB can reduce sustainment cost through three parallel mechanisms. The first is to reduce the need for additions to CHCS master files by enriching the comprehensiveness of the content. The TSB will provide standardized data sets to populate the CHCS master files. These data sets will be obtained from the mapped content of the DoD Reference Terminology for the domain (e.g. religion), which represents the combination of the concepts used by all 101 MTFs. To ensure compatibility with CHCS data, existing content is not replaced, but additional content is loaded.

Second, NCIDs are added into each site master file, using the mapping already completed in the HDD. NCIDs can then be sent as interface codes in transactions to CHCSII and any other ancillary systems, reducing the need for interface code mapping. With this approach, it no longer matters that the same item is assigned different IENs in different MTFs. An additional advantage is that different entries that mean the same thing will already have the same NCID. The master file will thus provide a rich set of representations (synonyms) to reduce the need to request a new entry, while maintaining concept normalization. There may be other benefits: Currently, when IENs or other CHCS legacy codes are used as interface codes, each MTF is set up with four interface contexts (Encounter/Demographics, Laboratory/Microbiology, Pharmacy and Text). Using NCIDs as interface codes may allow CHCSII configuration to be simplified at each MTF. Interface codes are checked against the HDD for every transaction, to translate to NCIDs for storage in the CDR. Sending NCIDs "directly" in transactions would eliminate the need for contexts and translations, which may enhance system performance.

The third mechanism to reduce sustainment cost is to manage deltas proactively. If a new entry is truly needed by one MTF, its addition will be propagated to all MTFs in the next master file update. Because updates are applied across all MTFs, the master file content will grow consistently and prospectively, reducing the need to create the same entry in other MTFs in an ad-hoc manner. The TSB is thus moving away from the current retrospective response to individual deltas to a proactive central management of DoD's vocabulary needs.

Operationally, the TSB will work with DoD to offer options for implementation that will meet DoD's workflow requirements. An example would be a tool to provide appropriate picklists from the DoD Reference Terminology, for selection of NCIDs to be added into site master files.