

May 2007

## Information Systems and Healthcare XVII: A HL7v3-based Mediating Schema Approach to Data Transfer between Heterogeneous Health Care Systems

Amar Gupta

*University of Arizona, [gupta@eller.arizona.edu](mailto:gupta@eller.arizona.edu)*

Richard Martin

*University of Arizona*

Prithi Avanavadi

*University of Arizona*

Surendra Sarnikar

*Dakota State University*

Follow this and additional works at: <https://aisel.aisnet.org/cais>

---

### Recommended Citation

Gupta, Amar; Martin, Richard; Avanavadi, Prithi; and Sarnikar, Surendra (2007) "Information Systems and Healthcare XVII: A HL7v3-based Mediating Schema Approach to Data Transfer between Heterogeneous Health Care Systems," *Communications of the Association for Information Systems*: Vol. 19 , Article 28.

DOI: 10.17705/1CAIS.01928

Available at: <https://aisel.aisnet.org/cais/vol19/iss1/28>

This material is brought to you by the AIS Journals at AIS Electronic Library (AISeL). It has been accepted for inclusion in Communications of the Association for Information Systems by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact [elibrary@aisnet.org](mailto:elibrary@aisnet.org).



**C**ommunications of the **I**nnovations for **S**ystems  
**A**ssociation for **I**nformation **S**ystems

## INFORMATION SYSTEMS AND HEALTHCARE XVII: A HL7V3-BASED MEDIATING SCHEMA APPROACH TO DATA TRANSFER BETWEEN HETEROGENEOUS HEALTH CARE SYSTEMS

Amar Gupta  
 Eller College of Management  
 University of Arizona  
[gupta@eller.arizona.edu](mailto:gupta@eller.arizona.edu)

Richard Martin  
 Prithi Avanavadi  
 University of Arizona

Surendra Sarnikar  
 Dakota State University

### ABSTRACT

One of the main challenges of exchanging patient care records between heterogeneous systems is the difficulty in overcoming semantic differences between them. This is further exacerbated by the lack of standardization in messaging protocols. As a solution to this problem, multiple ideas and standards have been proposed for exchanging clinical and administrative data in the healthcare area. However, most of these methods place some restrictions on the platform, standard or format, of the data. This paper proposes a context-specific, mediating schema-based architecture that enhances the transfer of electronic patient care records between healthcare information systems by using a reusable and portable model. The main contribution of this approach is its adaptability to a variety of schemas for the source and target systems.

**Keywords:** mediating schema, electronic patient care records, heterogeneous hospital systems, information exchange, HL7 version 3

### I. INTRODUCTION

Advances in Internet communication and database technology have made it possible to create and maintain patient information online [Shortliffe 1998]. However, the presence of multiple technology providers and varying healthcare agency policies has been a roadblock to the creation of a national integrated healthcare model. To some extent, regional healthcare models have been set up [Halamaka 2005] but are not scalable to higher levels. Standardization in messaging protocols in the healthcare area is the key to enabling interoperability between various healthcare systems and to allow electronic exchange of information between healthcare agencies. Several versions of messaging and information exchange standards have been published over the past decade. Toward this goal, the Health Level 7 (HL7) Working Group, under the aegis of the American National Standards Institute, has been involved in developing messaging and

information exchange standards for clinical and administrative data, in the healthcare domain [http://www.hl7.org]. Over the past few years, HL7 has undertaken a major initiative to design a new version of the messaging standard, HL7 version 3, which is the latest evolving standard for communication between the disparate information technology platforms of healthcare organizations. The new version was developed using an object-oriented development methodology and includes a reference information model at its core. The Health Level 7 version 3 includes several new features that include testability and conformance. The standard includes information models with several layers of abstraction to enable flexibility in creating messages and representing a large number of scenarios.

We believe HL7v3 is a very good candidate for a messaging standard in a national healthcare information network based on the following facts:

- It has an advanced messaging system that is already in place.
- It spans the entire healthcare realm rather than few specialized departments/procedures.
- There is significant amount of documentation available on the messaging system and past implementations.

In this paper, we demonstrate a mechanism for generating HL7v3 messages by reusing mappings between relational schemas and segments of HL7v3 specification that are common across different message specifications. Specifically, we present a mediating schema-based architecture and a procedure for generating HL7v3 messages by utilizing the reusable maps between the source schemas and Common Message Element Types (CMETs) of the HL7 specification. This mechanism enables the faster generation of HL7v3 messages and vastly reduces the complexity involved in generating HL7v3-compliant messages.

This paper is structured as follows. Section 2 describes schema mediation in the context of heterogeneous healthcare information systems, along with a review of related tools for generating schema maps. In Section 3, we provide an overview of the HL7 version 3 specifications and describe key components of the messaging architecture. We then present our proposed approach for generating HL7v3 messages and demonstrate the technique with an example in Section 4. This section also contains a case study. Finally, we conclude this paper with a discussion of the applicability of the proposed process in Section 5.

## II. BACKGROUND INFORMATION

We begin this section with a discussion of the interoperability of heterogeneous hospital systems and how schema mediation techniques can make the process of data integration and exchange more effective. We then present an overview of related efforts in the field along with a look at few commercial products.

### DATA EXCHANGE BETWEEN HETEROGENEOUS HEALTHCARE SYSTEMS

In a large decentralized healthcare network like that of the U.S., the variety and the number of healthcare information systems (HIS) are substantial. Disparate systems have difficulty communicating with one another because the data contained in each are structured differently, with different hierarchies and categorizations, data types, and definitions. Data exchange between heterogeneous systems can be enabled by developing a map between the source schema and the target schema [Wiederhold 1992]. Based on the classification of schema matching techniques by Rahm and Bernstein [2001], the linguistic matching technique is considered to be the most suitable for the healthcare arena as it gathers information, based on the names of tables and attributes in the source schemas, in a manner that does not violate any patient-privacy issues [Sarnikar and Gupta 2007].

The key to simplifying the complex nature of multiple client schema integration is the restriction of the domain. Without the specialized context of healthcare systems, the development of mediating

schemas would be inefficient and impractical. Furthermore, rather than exchanging an entire patient record, only data that are relevant to the type of specialized service provided at the data destination (called the context) are exchanged. Some examples of contexts within a healthcare system are pre-hospital care, clinical labs, and specialty clinics, such as ophthalmology and orthopedics. In creating a mediating schema, each context is studied to extrapolate certain reusable mappings from each of the client schemas.

## RELATED WORK

Significant research work in the area of data integration and interoperability between heterogeneous data sources has been and is currently taking place. Some of the projects can be directly applied to the healthcare setting, while others are generic projects. A few important projects and initiatives are discussed as follows.

### Context Interchange Framework (COIN)

The Context Interchange strategy [Sciore et al. 1994; Gupta and Madnick 1995] was developed as a non-intrusive, scalable, and accessible framework aimed at addressing the issues related to mediated data access from heterogeneous data sources. The framework represents knowledge from the data sources along with their contexts in an object-oriented model. The contexts (semantic metadata associated with both information sources and receivers) are collections of statements defining how data should be interpreted and how potential conflicts (differences in the interpretation) may be resolved. A general purpose context mediation service determines, according to the statements in the different contexts involved, what information is needed to answer the query, and generates a mediated query to reconcile the conflicts, which is then passed on to an optimizing query processor that executes the query.

### PANGAEA

PANGAEA [Maldonado et al. 2003] was developed as a middleware between applications and the underlying heterogeneous healthcare data sources. In this system, the integration approach is built on the concept of an archetype, which is a domain specific information structure based on a reference model. In this case, the reference model used is a European pre-standard called ENV13606. The archetypes form a semantic layer over the underlying data sources associating them with domain specific semantics through the use of a set of attribute matching functions between the archetypes and the schemas. These functions are also responsible for data translation between various data formats. A metadata server, which basically contains all the data definitions for the system, the archetypes-databases schemas mapping, and patients' social-demographic data, communicates with an Electronic Health Record server, which retrieves relevant patient information based on a user-given query.

### Integration Broker for Heterogeneous Information Sources (IBHIS)

The IBHIS project [Kotsiopoulos et al. 2004] aims to provide users with a unified view of data from heterogeneous sources on the fly, using a Service Oriented Data Integration Architecture (SODIA). Data sources are published as data services into a registry along with metadata and ontologies and are discovered on demand based on a given user query. The broker service is then invoked, which binds the data sources appropriately. An ontology service describes terms and services in the system to relate them semantically. Thus, the result is an integrated schema which is presented to the user.

### Integrating the Healthcare Enterprise (IHE) Initiative

IHE [[http://www.himss.org/ASP/topics\\_ihe.asp](http://www.himss.org/ASP/topics_ihe.asp)] is an initiative by healthcare professionals and organizations to promote the use of standards such as DICOM and HL7 for use in maintaining and exchanging health records. The framework defined by IHE consists of two parts – “Integration” profiles model the business process problem and the solution to the problem; and “Transactions” provide in-depth detail of how standards can be used to solve the problem mentioned. Once a problem (profile) is identified by professionals, vendors implement solutions to

this problem. These are then showcased, tested, and documented for future adoption by healthcare agencies.

Other notable efforts include The TSIMMIS project a Stanford University [Garcia-Molina et al. 1997] and the Health Data Integration (HDI) initiative at the E-Health Research Institute [<http://e-hrc.net/hdi/index.html>]. Besides these, there are a number of open source HL7 messaging tools present, some of which can be found at <http://www.hl7.org.au/HL7-Tools.htm>.

### COMMERCIAL SOLUTIONS

Several commercially available software solutions exist to partially mitigate the issue of transferring patient data from one system to another. The goal of many of these software solutions is to match patient data between healthcare sources such as ambulatory and hospital encounter data, pharmacy claims records, and laboratory test results [Jones and Sujansky 2004]. No standard method of identifying individual patients exists among healthcare providers; therefore associating patients across two or systems may rely on fuzzy matching techniques as in the case of LinkageWiz, SureMatch, DataSet V Suite, DeDupe4Excel, and other software products. These products vary in price and scope in order to suit varying need and have different strengths and weaknesses. In order to perform the matching, the data must be loaded into a relational database upon which scripts built into the software match the data based on product-specific algorithms. In general, the steps taken by all the software are to import the data, adjust the data to perform a field-by-field comparison, specify match weights, run the matching algorithms, display the actual and possible matches for manual review, and finally export the matched records. However, in order to perform the matches successfully, certain criteria need to be present, such as: basic understanding of probabilistic and fuzzy matching techniques, the availability of the data in tabular format, the existence of a master patient file, familiarity with the values of demographic fields in data, and the ability to post-process the matched results. The study by Jones and Sujansky [2004] does not discuss whether these products will match data that fulfill any of the schema incompatibilities listed in the study by Reddy et al. [1994].

In addition to the products mentioned previously, which perform generic schema mediation, there are HL7 specific products as well. They provide middleware solutions that enable the building, parsing, and exchange of HL7 messages between hospital systems. However, the disadvantage with these products is their restrictive nature when it comes to defining the client schema formats. For example, some products require the source/target schemas to be explicitly in a relational database.

## III. HL7v3 FRAMEWORK

### ARCHITECTURE OVERVIEW

In this section, we present a brief overview of the HL7 version 3 standard [Beeler 1998; Dolin 2006; Shakir 1997; Shakir 2003] and the major components of the message-building framework. The HL7v3 framework is a set of specifications that allow loosely coupled systems to interoperate. The core component of the framework is a reference information model (RIM). The RIM is a comprehensive object-oriented representation of the clinical, administrative, and financial aspects of the healthcare information systems. In other words, the RIM is a conceptual model of the healthcare domain. It is defined using a class diagram and contains information on the attributes, constraints, and associations between the classes [[http://www.miforum.net/distillate/rim/RIM0112\\_body.htm](http://www.miforum.net/distillate/rim/RIM0112_body.htm)]. All the HL7v3 message specifications are derived from the reference information model. A high level overview of the HL7v3 message building framework is shown in Figure 1.

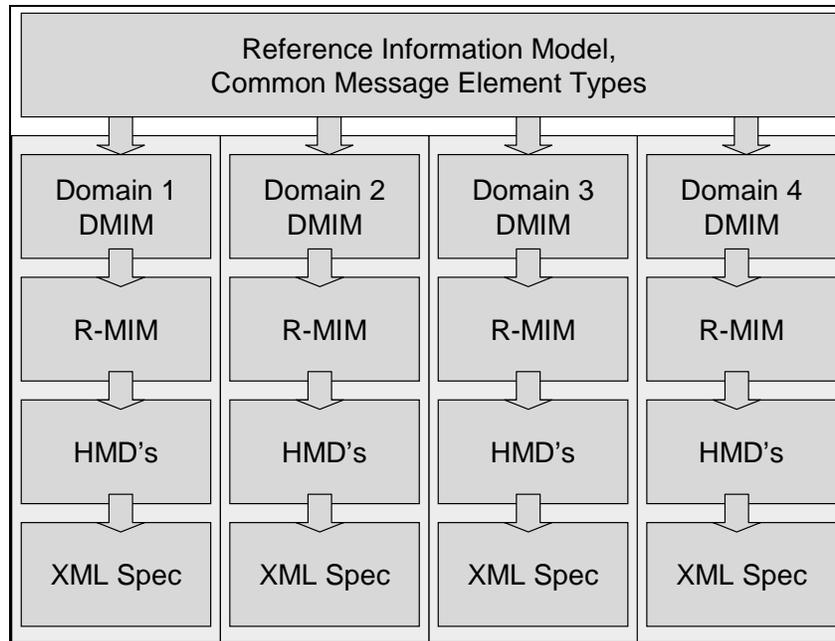


Figure 1. An Overview of the HL7v3 Message Building Framework (Adopted from Shakir 2003)

The reference information model can be further specialized into domain message information model (DMIM). The domain message information model is a subset of the reference information model and includes additional classes, attributes, and relationships that are specific to a particular domain area within the healthcare domain. Each message within a domain message information model is described in further detail in a refined message information model (R-MIM). The RMIM is a subset of the DMIM and describes the information content for a specific message.

The refined message information model is then converted into a hierarchical message definition. The hierarchical message definition consists of a tabular representation of the classes, attributes, and associations defined in a refined message information model. The hierarchical message definition is technology neutral and can be implemented in multiple technology standards as defined by an Implementable Technology Specification (ITS).

In addition to the previously described process for developing an XML specification, sections of the reference information model that are commonly used across domains models are abstracted into message fragments that can be reused across the derived domain specific message specifications. These message fragments are called the common message element types (CMET). The common message element types define message specifications for concepts that are common across several domains. Examples of such common concepts include Person, Address, and Organization.

### ROLE OF HL7V3 MESSAGES

If we attempt to share information directly between large numbers of healthcare data entities, a mapping must be created from each of the  $m$  sources to each of the  $n$  receivers, resulting in  $m \times n$  transformations. By using a mediating schema as an additional layer in between the sources and receivers the number of transformations can be limited to  $m + n$  [Sarnikar and Gupta,2007]. In order to achieve this, the mediating schema must be of a generic nature so that any client schema can be used. Also, the schema should provide an integrated collection of relevant attributes and metadata for patient care-related information. This is precisely where HL7v3 comes into play. As mentioned before, the HL7v3 is a comprehensive set of specifications that are developed for the entire healthcare domain. The use of this standard can allow healthcare data to be efficiently organized. Accordingly, a variety of data entities (such as Pre-hospital systems,

Emergency Department Systems, Laboratory Systems) can all be configured to parse and exchange HL7v3 messages, as shown in Figure 2.

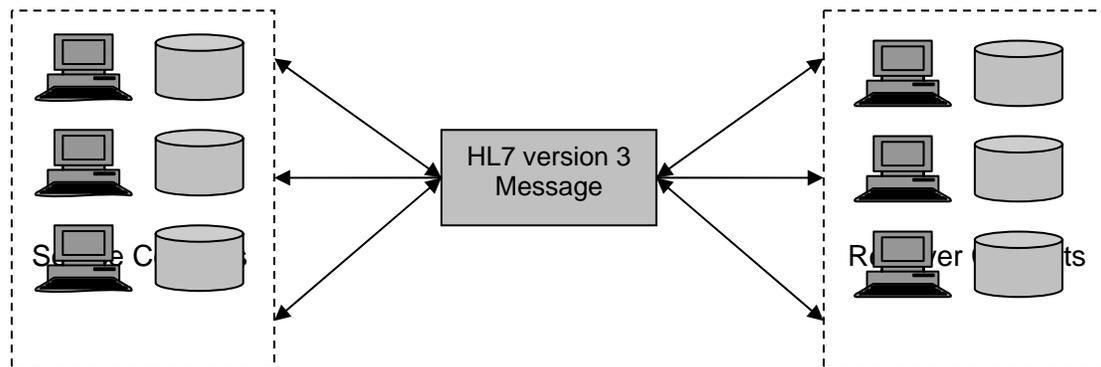


Figure 2. Architecture of System for Transferring Heterogeneous Data

While no algorithm could be developed to automate the process of schema matching [Sarnikar and Gupta 2007] due to the size and complexity of the schemas themselves, the strength of the proposed method lies in its reusability and its portability. Once a mapping has been generated between a source schema and the mediating schema, it can be reused until the original schema is altered. Furthermore, the data needs of disparate healthcare units almost always differ from each other. For example, the records of an EMT (Paramedic) unit might concentrate more on the type of incident and environment in which the medical service was administered, whereas the records of a hospital's emergency department might not. In such a case, the modular approach of HL7v3 will permit the selection of data that is relevant to the receiving context for transmission rather than the entire patient record. Therefore, a client need only store the mappings that are pertinent to its corresponding schema, making the system more scalable and portable.

Implementing organizations can extract several benefits from this latest messaging standard. However, given the size and complexity of the new standard, it involves a major development effort and software upgrade costs for participating organizations. Most data in organizations are stored in large relational databases, whose schema reflect the context-specific and customized information needs of the specific healthcare organization. The HL7 version 3 currently consists of over 100 hierarchical message descriptors, 1,300 fields, and 53 different data types and includes several dependencies among the message descriptors. Generating HL7v3 messages involves mapping relational data to hierarchical message descriptors, ensuring compliance with HL7v3 data standards, and conformance to rules embedded in the message specification. In addition, since the HL7v3 is an evolving standard, the relational data fields may need to be remapped to comply with revised versions of message specification.

#### IV. OUR APPROACH

In this paper, we build upon previous work in the area of schema mediation between heterogeneous healthcare information systems [Sarnikar and Gupta, 2007; Hass et al. 1997] by narrowing the scope to deal specifically with HL7v3 specifications as the mediating schema. We begin this section with a general overview of the process of heterogeneous data integration and exchange using context specific mediating schemas. We then describe our approach of using HL7v3 messages as the mediating schema between the source and target schemas.

#### GENERIC SCHEMA MEDIATION

As mentioned earlier, there have been several research efforts in the area of Schema Mediation [Gupta 1989; Wiederhold 1992; Bright et al. 1994; Milo and Zohar 1998; Shaker et al. 2002]. A generic methodology used in context specific schema mediation between healthcare information systems is outlined as follows.

1. A comparison is made between the source schema, the destination schema and the mediating schema structure. A mapping is generated between similar fields in the client schemas (source and destination) and the fields in the mediating schema. It is pertinent to note that the formats of the client schemas could range from comma-separated-value documents to relational database tables. As we are dealing with the specialized context of the healthcare industry, it is common to find certain similarities in the information captured within various client schemas. As mentioned before, the linguistic technique, which compares the attribute and table names to generate schema maps is used. As an example of this process, we consider the patient identifier, which will be the central entity in all healthcare information systems. Most frequently, a combination of fields such as Name, Address, SSN, or Driver's License will be used. Thus, the units (entities or attributes) used as patient identifiers are easily associated across different units and thus, a mapping can be created.
2. Once the mappings are generated between the client schemas and the mediating schema, the patient care record (PCR) containing the relevant information from the source schema is generated. Since the source and receiver contexts are known, any general querying method, such as SQL for relational tables can be used to obtain the desired fields from the client schemas. The patient care record from the source, obtained as a result of the query is converted to an XML document. Therefore, this step generates two XML documents corresponding to the source PCR and a target placeholder PCR. This entire process was described by Popa et al. [2002].
3. The source and target XML files, along with the mediating schema structure are used to transform the two client-to-mediator mappings to a unique source-to-target mapping, which can be reused for data translation between source and target at any time. Thus, the final output of the process is a target PCR containing the information required to populate the target schema.

### HL7V3 IN SCHEMA MEDIATION

Our approach of using HL7v3 as the mediating schema follows the same procedure as the one outlined in the previous paragraphs. The advantage of the HL7v3 messaging system is that the initial step of generating mappings between the client schemas and the mediating schema is greatly simplified due to the Common Message Element Types (CMET) present in the specifications. As mentioned earlier, a CMET is a block of XML that is common across different HL7 messages. Some HL7 entities designated as CMETs (there are more than a hundred) will almost certainly have as corresponding entities in the client schema, depending on the particular healthcare domain being considered. Thus, by manual inspection or semi-automated procedures of text comparison between the attribute/entity names in the client and mediating schemas, mappings are generated and stored. For example, the field(s) used to store the address of the patient in the source schema is mapped to the address CMET (named "Address") in the HL7 specification. Most probably, the name of the field in the source schema will also be similar (if not equal) to "Address."

The next step is to analyze the hierarchical message definition (HMD) using the maps obtained from the previous step. As depicted in Figure 1, the HMD is very close to the final HL7 message. The HMD consists of CMETs along with other entities from the HL7 specifications. The CMETs present in the HMD are identified. Then the maps previously developed between the client schema and the HL7 specifications are used to generate the actual target message. The CMETs are mapped directly and the other units can be mapped using any of the schema matching techniques previously outlined in section 3. The process is depicted in Figure 3.

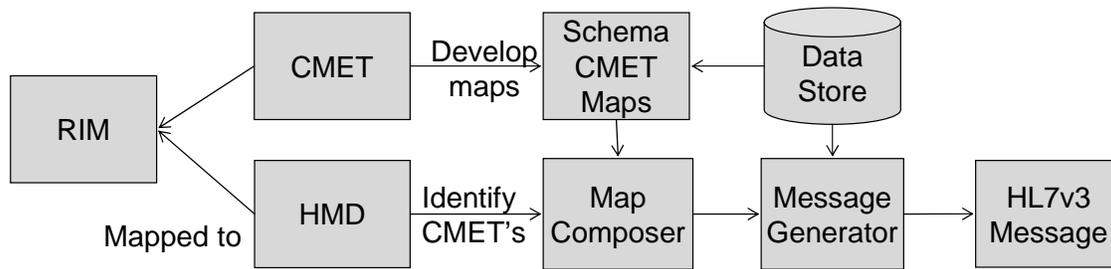


Figure 3. An Overview of the HL7v3 Message Generation Process

The following paragraphs contain a case study illustrating the concepts and approach that was outlined previously. As an operational scenario, we will consider the creation of an HL7 message for identifying a particular patient. For illustrative purposes, we assume that the source schema is in Comma-Separated-Value (CSV) format. The case study will cover step-by-step our proposed method generating the mapping between the source schema and the HL7 specification as this is the most important step in the data translation process. The main components involved in the process are as follows:

1. *Source CSV Structure file (SCS)* – This is an XML version of the source database containing blocks for each attribute.
2. *Source Data file (CSV)* – This is the actual CSV file containing the patient identifying information.
3. *HL7v3 Specification file (H3S)* – This is an XML file containing the description of the CMET used for identifying patients.
4. *CSV to HL7 Map file (MAP)* – The mappings created between attributes in the SCS file and the H3S file are stored in this file.
5. *HL7v3 Message file* – This is the final output of the process, containing the HL7 message to be transmitted.

A pictorial representation of the mediation process with respect to the above components is shown in Figure 4. This is followed by some example snippets from each of the intermediate files generated.

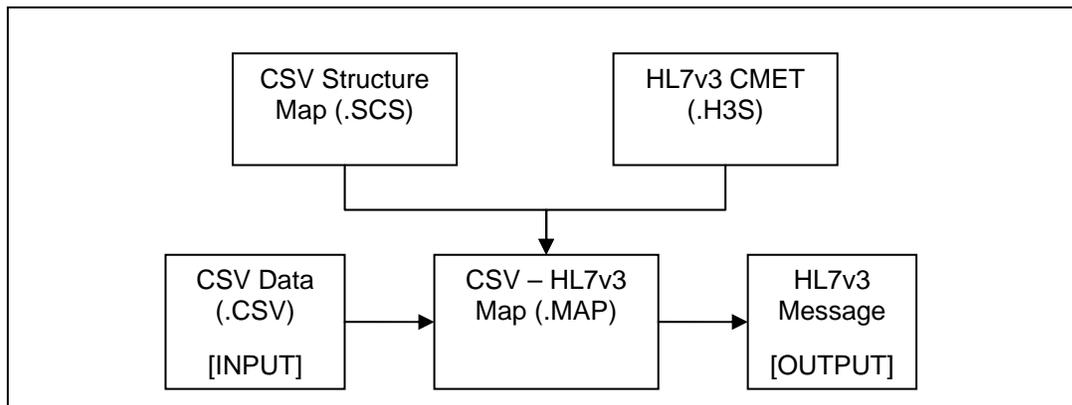


Figure 4. HL7v3 Mediation Process Overview

```

<?xml version="1.0" encoding="UTF-8"?>
<csvMetadata uuid="f9364790-7e79-4d6b-8ed7-fa2e7d158f15" version="1.2">
<segment name="INCIDENT_NUM" uuid="bd0d58a0-9d08-4bea-8ff4-636352d56ae2">
<field column="1" name="Carrier" uuid="e557501e-4e29-4d06-bd69-35738ed005ab"/>
<field column="2" name="Rig_Num" uuid="df25f96d-58e1-41bd-82e6-2feb4206969d"/>
<field column="3" name="Patient_Num" uuid="165b425e-996f-4a76-addc-999f34519517"/>
<field column="4" name="Patient_fname" uuid="e29d062b-1e3a-4896-8ea8-ec097a2c2c8d"/>
<field column="5" name="Patient_lname" uuid="345fbf47-de44-4dde-a051-44ff8a3607a5"/>
<field column="6" name="Patient_DOB" uuid="70f8949e-8a7e-47c5-9894-789da641b4c9"/>
<field column="7" name="Patient_Age" uuid="d589500b-d0ff-493c-85cf-43234fc6265c"/>
<field column="8" name="Patient_Sex" uuid="11d6f650-a74a-487b-b686-92d58700c106"/>
<field column="9" name="Patient_Phone" uuid="a8659027-71c3-4623-bd7f-8be502557384"/>
<field column="10" name="Patient_Addr1" uuid="3b0add4e-de13-40cf-9ed2-46b25c9ad8fd"/>
<field column="11" name="Patient_Addr2" uuid="85319a10-62f5-451c-9f07-e6998e3e7b18"/>
<field column="12" name="Patient_City" uuid="97cc542a-afbc-453d-a6ce-6950852879f9"/>
<field column="13" name="Patient_State" uuid="e9632401-eb43-4e18-84d9-3c2bd78e2d5f"/>
<field column="14" name="Patient_Zip" uuid="752e7cd3-b485-4bd6-99aa-5ab619a4ff79"/>
<field column="15" name="Patient_SSN" uuid="c2c54914-bcc5-42d8-b972-cef5b8110609"/>
<field column="16" name="Patient_InsureCo" uuid="2cb02112-6f66-4eea-a68c-e00b07457945"/>
<field column="17" name="Patient_InsrNum" uuid="a0f2178b-cceb-4c4a-836d-b39844ece89e"/>
</segment>
</csvMetadata>

```

Figure 5. Source CSV Specification File

```

INCIDENT_NUM, 1, Medflight, 7, 873479, John, Doe, 1-1-1926, 80, M, 520-555-5555, 456
Washington Blvd, Suite 1000, Washington, DC, 20002, 123-45-6789, United

```

Figure 6. Sample Data from the Source Database

```

<?xml version="1.0" encoding="UTF-8"?>
<hl7v3meta messageId="COCT_MT090102" version="V1.2">
<clone clonename="AssignedPerson" sortKey="1" cardinality="1..1" uuid="18b8a04b-6205-4661-
be41-c164a38ea7a9">
<attribute uuid="31e2c05f-1930-4dc7-a1ef-404f08978502" name="classCode" sortKey="1"
cardinality="1..1" isMandatory="true" conformance="R" codingStrength="CNE"
domainName="RoleClass" hl7-default="ASSIGNED" rimSource="Role"/>
<attribute uuid="32209d07-1183-43c6-b230-390a9ff4869f" name="id" sortKey="2"
cardinality="1..*" datatype="II" isMandatory="true" conformance="R" rimSource="Role"
multipleSequenceNumber="1">
<datatypeField name="root" uuid="966c5f15-02bc-48a1-994d-76b1867b3c41" cardinality="1..1"/>
<datatypeField name="extension" uuid="d104d3f3-7de6-406a-b0d6-45d5046271a9"
cardinality="0..1"/>
<datatypeField name="inlineText" uuid="70d32bcd-7e37-4d38-8b4d-a7dba1fd0812"
cardinality="0..1"/>
</attribute>

<attribute uuid="c83f28a2-ffd7-4cca-b5ee-60ce37e34fa7" name="county" sortKey="2"
cardinality="0..*" multipleSequenceNumber="1">
<datatypeField name="inlineText" uuid="e20066bf-956f-4946-8318-871443c4db00"
cardinality="0..1"/>
</attribute>
</clone>
</clone>
</hl7v3meta>

```

Figure 7. Section of the HL7 Specification File

### Source Schema Specification and HL7 CMET Specification

As explained previously, the source specification is maintained in the SCS file. This file contains metadata about the various fields used to capture patient identifying information in the source database. A sample SCS file is shown in Figure 5. A sample record containing the actual patient data from the source database is shown in Figure 6. This record contains values for all the attributes in the SCS file.

The CSV specification file is used to generate maps between the elements in the file at the CMETs in the HL7 specification. A section of the HL7 specification file for the CMET corresponding to patient identifying information is shown below in Figure 7. Because HL7 is a generic comprehensive standard, it attempts to capture large amounts of information for a particular CMET. Due to large file size, the sample shown is highly condensed.

### Source - CMET Mappings

One of the greatest challenges of transferring semantic information between different schemas is to avoid the potential loss of meaning within the data. Collecting and processing the meaning of

data is, as yet, not automatable and must be done through human medium. As such, the mappings between the client schema and the HL7 CMET Specification were developed manually and consist of mappings between corresponding elements. An example of client to mediator mappings is shown in Figure 8.

```
<?xml version="1.0" encoding="UTF-8"?>
<mapping version="1.2">
<components>
<component kind="scs" location="C:\Integrator\HL7Adaptor\iReviveCSVStructure.scs"
type="source" uuid="13d2fa67-252a-456a-b404-45365a7fef7d"/>
<component kind="HL7v3" location="C:\Integrator\HL7Adaptor\PatientIdentifyingInfo.h3s"
type="target" uuid="22e3d3c5-6503-48e8-ba09-7ceca0fea8b1"/>
</components>
<links>
<link uuid="103918ac-56fd-4b28-bddf-1fb821dd3040">
<linkpointer component-uuid="13d2fa67-252a-456a-b404-45365a7fef7d" data-uuid="165b425e-
996f-4a76-addc-999f34519517"/>
<linkpointer component-uuid="22e3d3c5-6503-48e8-ba09-7ceca0fea8b1" data-uuid="966c5f15-
02bc-48a1-994d-76b1867b3c41"/>
</link>
...
<link uuid="86cf37ce-9dcb-4523-b6c0-bca21b079034">
<linkpointer component-uuid="13d2fa67-252a-456a-b404-45365a7fef7d" data-uuid="df25f96d-
58e1-41bd-82e6-2feb4206969d"/>
<linkpointer component-uuid="22e3d3c5-6503-48e8-ba09-7ceca0fea8b1" data-uuid="d104d3f3-
7de6-406a-b0d6-45d5046271a9"/>
</link>
</links>
<views>
<view component-uuid="13d2fa67-252a-456a-b404-45365a7fef7d" height="0" width="0" x="0"
y="0"/>
<view component-uuid="22e3d3c5-6503-48e8-ba09-7ceca0fea8b1" height="0" width="0" x="0"
y="0"/>
</views>
</mapping>
```

Figure 8. A Sample Subsection of a CSV to HL7 Mappings

Once the mappings have been generated, a source PCR file, which is the actual HL7v3 message, is generated using the mappings generated and the actual source data records. This message is then sent across to the destination. At the destination, the mappings between HL7v3

CMETs and the target schema are analyzed. The actual source data from the message corresponding to the CMETs in the target mappings are extracted. The rest of the fields are mapped using any semi-automated process described in previous sections and in existing literature. An example HL7v3 message is shown in Figure 9.

```
<?xml version="1.0" encoding="UTF-8"?>
<COCT_MT090102.AssignedPerson xmlns="urn:hl7-org:v3"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" classCode="ASSIGNED">
<id xsi:type="II" root="7" extension="Medflight" assigningAuthorityName="1"/>
<addr xsi:type="AD">
<postalCode representation="TXT" mediaType="text/plain" partType="ZIP">20002
</postalCode>
<state representation="TXT" mediaType="text/plain" partType="STA">DC</state>
<streetAddressLine representation="TXT" mediaType="text/plain" partType="SAL">456
Washington Blvd </streetAddressLine>
<additionalLocator representation="TXT" mediaType="text/plain" partType="ADL"> Suite
1000</additionalLocator>
<city representation="TXT" mediaType="text/plain" partType="CTY">Washington</city>
</addr>
<person classCode="PSN" determinerCode="INSTANCE">
<name xsi:type="EN">
<family partType="FAM" representation="TXT" mediaType="text/plain">Doe</family>
<given partType="GIV" representation="TXT" mediaType="text/plain">John</given>
</name>
</person>
</COCT_MT090102.AssignedPerson>
```

Figure 9. A HL7v3 Message

## V. FUTURE POSSIBILITIES AND CONCLUSION

ERP or Enterprise Resource Systems are frequently used by companies that wish to streamline the supply chain and provide one-stop-shopping accessibility to information needed across departments. The use of ERP systems reduces the redundancy of independent data repositories and provides a meaningful connection between departmental silos. The idea of integrating departmental silos can also be applied to the healthcare field in various levels of granularity. For example, within one hospital, each specialized care service (pediatrics, orthopedics, emergency, etc) can represent different departments. With this concept in mind, some health care organizations have begun reorganizing in order to implement new technology that will integrate data silos such as hospital suppliers, hospital providers, physicians and specialists, and patients [Siau 2003]. The proposed method of data translation discussed in this paper aims to build a data link between heterogeneous patient-to-patient data systems but can be used to provide a data link between suppliers, providers, physicians, and patients via extranet or Internet.

To summarize, the issues relating to health data information integration and exchange as follows

- The diversity of hospital organizations, complexity of medical procedures and various preferences of healthcare organizations give rise to several islands of information containing patient data.
- The emergence of HL7v3 as a standard messaging protocol allows the adoption of a global schema for healthcare information.
- The unwillingness of healthcare organizations to alter their data due to competitive reasons, future bottlenecks, or the possible loss of critical data further complicates the issue.
- The sensitive nature of health records is a barrier to rapid data integration and exchange.
- Traditional solutions such as federated database systems and global data warehouses fail to take into account the constantly changing healthcare environment.

To circumvent the above issues, we have proposed a method to transfer electronic patient care records between heterogeneous hospital systems through a context specific mediating approach. When the mediating schema used is the HL7v3 specification, the translation process is made more effective and efficient. In contrast to most other approaches, our approach does not restrict the system to a particular platform or data model. The modular nature of the mediating schemas enables portability, sharing and standardization even in the presence of disparity among data sources. In future research, we intend to explore mechanisms for automated translation between different data coding schemes and mechanisms for integrating multiple context mediators.

#### ACKNOWLEDGEMENTS

The authors would like to acknowledge Steve Moulton, Mark Gaynor, and Dan Myung for their helpful comments and suggestions.

#### REFERENCES

- Beeler, G. (1998). "HL7 Version 3 — an Object-Oriented Methodology for Collaborative Standards Development," *International Journal of Medical Informatics*. (48)1-3, pp. 151-161.
- Bright M., A. Hurson, and S. Pakzad. (1994). "Automated Resolution of Semantic Heterogeneity in Multidatabases," *ACM Transactions on Database Systems*. (19)2, pp. 212–253.
- Dolin, R. et al. (2006). "The HL7 Clinical Document Architecture," Release 2, *Journal of the American Medical Informatics Association*. (8), pp. 552-569.
- Felleiter, P., M. Helm, L. Lampl, and K. Bock. (1995). "Data Processing in Pre-hospital Emergency Medicine," *International Journal of Clinical Monitoring and Computing*. (12) 1, pp. 37-41.
- Garcia-Molina, H., et al. (1997). "The TSIMMIS Approach to Mediation: Data Models and Languages," *Journal of Intelligent Information Systems*.
- Gupta, A. (ed.) (1989). *Integration of Information Systems: Bridging Heterogeneous Databases*. IEEE Press, New York, NY.
- Gupta, A. and S. E. Madnick. (1995). "Context Interchange: Research in Using Knowledge about Data to Integrate Disparate Sources," PROFIT Working Paper #95-29, MIT Sloan School of Management, Cambridge, MA.
- Haas, L., R. Miller, B. Niswonger, M. Roth, and E. Wimmers. (1997). "Transforming Heterogeneous Data with Database Middleware: Beyond Integration," *IEEE Data Engineering Bulletin*.

- Halamka, J., et al. (2005). "Healthcare IT Collaboration in Massachusetts: The Experience of Creating Regional Connectivity," *Journal of American Medical Information Association*. (12), 596-601.
- Jones, L. and W. Sujansky. (2004). "Patient Data Matching Software: A Buyer's Guide for the Budget Conscious," *iHealthReports*. August, pp. 1-28.
- Kotsiopoulos, I., J. Keane, M. Turner, P. Layzell, Z. Funjun. (2003). "IBHIS: Integration Broker for Heterogeneous Information Sources," Proceedings of the 27<sup>th</sup> Annual International Computer Software and Applications Conference, pp. 378-384.
- Larson, J., S. Navathe, and R. Elmasri. (1989). "A Theory of Attribute Equivalence in Databases with Application to Schema Integration," *IEEE Transactions on Software Engineering*. (16) 4, pp. 449-463.
- Maldonado, J. A., M. Robles, and P. Crespo. (2003). "Integration of Distributed Healthcare Records," Proceedings of the 16<sup>th</sup> IEEE Symposium on Computer-based Medical Systems, pp. 213-218.
- Milo, T. and S. Zohar. (1998). "Using Schema Matching to Simplify Heterogeneous Data Translation," Proceedings of the 24th International Conference On Very Large Data Bases, pp. 122-133.
- Popa, L., V. Velegakis, R. Miller, M. Hernandez, and R. Fagin. (2002). "Translating Web Data," Proceedings of the Very Large Databases Conference, Hong Kong, China.
- Rahm, E. and P. Bernstein. (2001). "A Survey of Approaches to Automatic Schema Matching," *Very Large Database Journal*. (10) 4, pp. 334-350.
- Reddy, M. and A. Gupta. (1995). "Context Interchange: A Lattice Based Approach," *Knowledge-Based Systems*. (8) 1.
- Reddy, M. P., B. E. Prasad, and A. Gupta. (1995). "Formulating Global Integrity Constraints during Derivation of Global Schema," *Data and Knowledge Engineering*. (16), pp. 241-268.
- Reddy, M. P., B. E. Prasad, P. G. Reddy, and A. Gupta. (1994). "A Methodology for Integration of Heterogeneous Databases," *IEEE Transactions on Knowledge and Data Engineering*. (6) 6.
- Sarnikar, S. and A. Gupta. (2007). "A Context-Specific Mediating Schema Approach for Information Exchange between Heterogeneous Hospital Systems," Forthcoming in *International Journal of Healthcare Technology and Management*.
- Sciore, E., M. Siegel, and A. Rosenthal. (1994). "Using Semantic Values to Facilitate Interoperability among Heterogeneous Information Systems," *ACM Transactions on Database Systems*. (19), pp. 254-290.
- Shakir, A., M., (1997). "HL7 Reference Information Model: More Robust and Stable Standards," *Healthcare Informatics*. (14) 7, p. 68.
- Shakir, A., (2003). "HL7 Version 3 Advanced Tutorial," Accessible at <http://www.hl7.org/library/committees/mnm/presentations/Version%203%20Advanced%20Tutorial%20%28W7%29%2Ezip>.
- Shaker, R., P. Mork, M. Barclay, and P. Tarczy-Hornoch. (2002). "A Rule Driven Bi-directional Translation System for Remapping Queries and Result Sets between a Mediated Schema and Heterogeneous Data Sources," *Proceedings of the AMIA Symposium*, pp. 692-696.
- Shortliffe, E. H. (1998). "The Evolution of Health-Care Records in the Era of the Internet," *Medinfo*. 9(suppl), pp. 8-14.

Siau, K. (2003). "Health Care Informatics," *IEEE Transactions on Information Technology in Biomedicine*. (7) 1, pp. 1-7.

Wiederhold, G. (1992). "Mediators in the Architecture of Future Information Systems," *IEEE Computer*. (25) 3, pp. 38-49.

#### ABOUT THE AUTHORS

**Amar Gupta** is a Tom Brown Endowed Chair of Management and Technology; Professor of Entrepreneurship, Management Information Systems, Management of Organizations, and Computer Science at the University of Arizona. Earlier, he was with the MIT Sloan School of Management (1979-2004); for half of this 25-year period, he served as the founding co-director of the Productivity from Information Technology (PROFIT) initiative. He has published over 100 papers, and serves as associate editor of *ACM Transactions on Internet Technology*. He has nurtured the development of several key technologies that are in widespread use today and is currently focusing on the area of the 24-Hour Knowledge Factory.

**Richard Martin** is a Masters student in the Department of Computer Science at the University of Arizona. His areas of interest are databases and Internet technologies.

**Prithi Avanavadi** has graduated with a Masters degree in Management Information Systems from the University of Arizona. She is currently working at Raytheon Corporation.

**Surendra Sarnikar** is currently an assistant professor in Information Systems at the College of Business and Information Systems at Dakota State University. His work focuses on knowledge management and technologies for knowledge dissemination and discovery. Earlier, he obtained his Ph.D in Management Information Systems from the University of Arizona.

Copyright © 2007 by the Association for Information Systems. Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and full citation on the first page. Copyright for components of this work owned by others than the Association for Information Systems must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, or to redistribute to lists requires prior specific permission and/or fee. Request permission to publish from: AIS Administrative Office, P.O. Box 2712 Atlanta, GA, 30301-2712 Attn: Reprints or via e-mail from [ais@aisnet.org](mailto:ais@aisnet.org)



# Communications of the Association for Information Systems

ISSN: 1529-3181

## EDITOR-IN-CHIEF

Joey F. George  
Florida State University

## AIS SENIOR EDITORIAL BOARD

Jane Webster Vice President Publications Queen's University	Joey F. George Editor, CAIS Florida State University	Kalle Lyytinen Editor, JAIS Case Western Reserve University
Edward A. Stohr Editor-at-Large Stevens Inst. of Technology	Blake Ives Editor, Electronic Publications University of Houston	Paul Gray Founding Editor, CAIS Claremont Graduate University

## CAIS ADVISORY BOARD

Gordon Davis University of Minnesota	Ken Kraemer Univ. of Calif. at Irvine	M. Lynne Markus Bentley College	Richard Mason Southern Methodist Univ.
Jay Nunamaker University of Arizona	Henk Sol Delft University	Ralph Sprague University of Hawaii	Hugh J. Watson University of Georgia

## CAIS SENIOR EDITORS

Steve Alter U. of San Francisco	Jane Fedorowicz Bentley College	Chris Holland Manchester Bus. School	Jerry Luftman Stevens Inst. of Tech.
------------------------------------	------------------------------------	---	---

## CAIS EDITORIAL BOARD

Michel Avital Univ of Amsterdam	Erran Carmel American University	Fred Davis Uof Arkansas, Fayetteville	Gurpreet Dhillon Virginia Commonwealth U
Evan Duggan Univ of the West Indies	Ali Farhoomand University of Hong Kong	Robert L. Glass Computing Trends	Sy Goodman Ga. Inst. of Technology
Ake Gronlund University of Umea	Ruth Guthrie California State Univ.	Alan Hevner Univ. of South Florida	Juhani Iivari Univ. of Oulu
K.D. Joshi Washington St Univ.	Michel Kalika U. of Paris Dauphine	Jae-Nam Lee Korea University	Claudia Loebbecke University of Cologne
Paul Benjamin Lowry Brigham Young Univ.	Sal March Vanderbilt University	Don McCubbrey University of Denver	Michael Myers University of Auckland
Fred Niederman St. Louis University	Shan Ling Pan Natl. U. of Singapore	Kelley Rainer Auburn University	Paul Tallon Boston College
Thompson Teo Natl. U. of Singapore	Craig Tyrant W Washington Univ.	Chelley Vician Michigan Tech Univ.	Rolf Wigand U. Arkansas, Little Rock
Vance Wilson U. Wisconsin, Milwaukee	Peter Wolcott U. of Nebraska-Omaha	Ping Zhang Syracuse University	

## DEPARTMENTS

Global Diffusion of the Internet. Editors: Peter Wolcott and Sy Goodman	Information Technology and Systems. Editors: Alan Hevner and Sal March
Papers in French Editor: Michel Kalika	Information Systems and Healthcare Editor: Vance Wilson

## ADMINISTRATIVE PERSONNEL

Eph McLean AIS, Executive Director Georgia State University	Chris Furner CAIS Managing Editor Florida State Univ.	Copyediting by Carlisle Publishing Services
---	---	--