

USARIEM TECHNICAL REPORT T- 02/12

**AUTOMATED DATA MANAGEMENT FOR
WARFIGHTER PHYSIOLOGIC STATUS MONITORING**

Mark J. Buller¹
Rob M. Siegel¹
Gary P. Vailllette¹
Debra Meyers¹
William T. Matthew²
Stephen P. Mullen²
Reed W. Hoyt²

¹GEO-CENTERS INC.
7 Wells Avenue
Newton, MA

²Biophysics and Biomedical Modeling Division

February 2002

DISTRIBUTION STATEMENT A
Approved for Public Release
Distribution Unlimited

U.S. Army Research Institute of Environmental Medicine
Natick, MA 01760-5007

20020204 034

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE DEC 2001	3. REPORT TYPE AND DATES COVERED TECHNICAL REPORT	
4. TITLE AND SUBTITLE AUTOMATED DATA MANAGEMENT FOR WARFIGHTER PHYSIOLOGIC STATUS MONITORING		5. FUNDING NUMBERS	
6. AUTHOR(S) M. BULLER, R. SIEGEL, G. VAILLETTE, D. MEYERS, W. MATTHEW, S. MULLEN, AND R. HOYT		8. PERFORMING ORGANIZATION REPORT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) U.S. Army Research Institute of Environmental Medicine Kansas Street Natick, MA 01760-50076		10. SPONSORING / MONITORING AGENCY REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Materiel Command Fort Detrick, MD 21702-5007		11. SUPPLEMENTARY NOTES	
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution unlimited		12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words) The Warfighter Physiological Status Monitoring (WPSM) program produces millions of data points per field study. The effective utilization of this growing collection of very large time series data sets is crucial for achieving the scientific goals of the WPSM program. Currently available object-relational database systems do not deal well with either temporal or time series data, making the development of a suitable system costly, time consuming, and difficult. A highly automated data management solution is described here which allows investigators rapid and easy access to pertinent and interesting subsets of data. The approach uses object-oriented methods and the Extensible Markup Language (XML), a World Wide Web (WWW) standard for information exchange, to standardize field study data in an extensible way. Rather than using a full-scale object-relational database management system, an XML file archive was developed. A client-server based application was engineered to provide a generic interface to the WPSM XML data archive, and to provide a simple to use graphical user interface. The complete archive and system of software was used in four different applications and proved successful in its goal of automating the collection, archiving and access of data.			
14. SUBJECT TERMS Time Series Data Management; Temporal Data; Data Automation; Data Access; XML Extensible Markup Language; Object-Oriented; Object-Relational; Database Management Systems; DBMS; Data Mining; Data Viewing; Bio-Informatics; Medical Informatics		15. NUMBER OF PAGES 77	16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT U	18. SECURITY CLASSIFICATION OF THIS PAGE U	19. SECURITY CLASSIFICATION OF ABSTRACT U	20. LIMITATION OF ABSTRACT U

TABLE OF CONTENTS

<u>SECTION</u>	<u>PAGE</u>
LIST OF FIGURES	vi
LIST OF TABLES	vii
EXECUTIVE SUMMARY	1
INTRODUCTION	2
DATA VOLUME PROBLEM.....	2
BENEFITS OF AUTOMATION.....	2
METHODS.....	3
DATA CHARACTERIZATION.....	3
Data Standardization.....	4
DATA ARCHIVE	5
Two Alternate Approaches.....	6
WPSM Data Archive	7
Context and Revision Control.....	7
DATA ACCESS.....	8
WPSM XML Query Server	9
WPSM User Interface Client	9
RESULTS.....	11
U.S. MARINE CORPS INFANTRY OFFICER COURSE (SEPTEMBER 1999) ..	11
U.S. MARINE CORPS INFANTRY OFFICER COURSE (JULY 2001)	12
U.S. CENTER FOR ENVIRONMENTAL HEALTH RESEARCH (USACEHR)	12
SCENARIO-J	12
DISCUSSION	13
LIMITATIONS	14
ETHICAL USE OF ONLINE DATA.....	14
CONCLUSIONS	14
REFERENCES	15
APPENDIX A. LITERATURE REVIEW.....	18
METHODS	18
Database Searches and Keywords	18

<u>SECTION</u>	<u>PAGE</u>
List of Keywords used in Literature Search	20
Projects and Products	21
Research Institutes, Associations, and Societies	23
 APPENDIX B. WPSM XML ENCODED DATA	 25
BACKGROUND	25
SOME BASIC XML	25
DATA OBJECT DEFINITION	26
Object Data Type	27
Persistence	27
Time	28
ID	29
Spatial Elements	29
User Defined Elements	29
ASSIGNMENT OF DATA TYPES	30
Implicit Assignment	30
Explicit Assignment	30
WPSM STUDY ARCHIVE.....	31
Directory Structure	31
Directory Naming Convention	32
XML File Naming Convention.....	32
Other Directories in the Archive	33
 APPENDIX C. WPSM XML DATA DICTIONARY	 34
WPSM XML DATA TYPES	34
DATA TYPE ELEMENT DESCRIPTIONS	35
 APPENDIX D. WPSM XML DATA CODING SCHEMES	 45
ACTIVITY CODING SCHEME	45
CLOTHING CODING SCHEME.....	53
EQUIPMENT CODING SCHEME	56
FOOD TABLE	57
 APPENDIX E. WPSM QUERY SERVER TEXT INTERFACE STANDARD.....	 59
GQLCommand.txt	59
GQLQueryString.txt	59
GQLStudySummary.txt	60
GQLQueryResults.txt	61
GQLQueryResults.bin.....	62

SECTION

PAGE

PROCESS OF QUERYING A WPSM XML ARCHIVE USING THE QUERY
SERVER 62

BIBLIOGRAPHY 63

LIST OF FIGURES

<u>Figure</u>		<u>Page</u>
1	WPSM XML data example.....	5
2	Database model selection based upon data and query complexity	6
3	WPSM XML sample data archive structure.....	7
4	Block functional diagram of data access application architecture...	8
5	User interface showing data set, data type selection.....	10
6	Graphical display of query results.....	11
7	Block functional diagram of WPSM data management architecture with link to SCENARIO Model.....	13
B1	Complete WPSM XML Data File.....	26
B2	Fully Defined Data Object.....	26
B3	Implicit Assignment.....	30
B4	Explicit Assignment.....	31
B5	WPSM Study Archive Directory Structure.....	31

LIST OF TABLES

<u>Table</u>		<u>Page</u>
1	Typical WPSM Field Study Data.....	3
C1	WPSM Data Types, Persistence, and Entities.....	34
D1	Reported and Coded Activities.....	48
D2	Reported and Coded Clothing.....	54
D3	Reported and Coded Equipment.....	56
D4	Food Table.....	57

EXECUTIVE SUMMARY

The Warfighter Physiologic Status Monitoring (WPSM) program uses state of the art ambulatory monitoring technologies to track and measure subject status over many days. With the use of this technology, an average field study will yield over 1.5 million data points--deluging the researcher. Data on this scale means that research is limited to the ability to process and analyze the raw data, rather than by the basic science. To overcome this data management problem, we used a standard three-step approach to managing large data sets. That is, (1) characterize, (2) archive, and (3) retrieve data. First, an extensible object-oriented based schema for characterizing WPSM data was implemented. These schema are captured and encoded using the Extensible Markup Language (XML), a standard for business-to-business information exchange across the world wide web (WWW). Second, current and emerging database technologies were explored as a means for archiving WPSM data. Extant relational and object-relational databases are not designed to manage temporal and time series data. Since current database technology tended to be costly, time-consuming, and unable to provide quick relief from data overload, a second approach was selected where simple directory structures and an XML file archive was used. Hypertext Markup Language (HTML) documents provided context and notes for data revisions. Third, access to the data was provided by custom client-server software. Two software components provided both a generic software interface to the data and a simple, easy to use graphical user interface (GUI). The two components work in concert to provide rapid access to data for working scientists. For example, data from the U.S. Marine Corps Infantry Officer Course (September 1999) were the first to be represented in XML. The system's ability to time-correlate and provide quick access to subsets of data for further analysis demonstrated the utility of the data automation architecture. The automation process was tested in full during the U.S. Marine Corps Infantry Officer Course (July 2001), where data were downloaded from individual loggers and converted to XML. During this study, data were available almost instantaneously for viewing and analysis. This approach has also been used to manage large amounts of biosensor data in a U.S. Army Center for Environmental Health Research project, and has been adapted to provide automated data feeds to predictive models, allowing the simultaneous viewing of real versus predicted results. The standardized field study XML data format and the data viewing mining tools have proven useful to the WPSM program, enabling the automation of the data collection, data archiving, and data accessing process.

INTRODUCTION

New ambulatory physiological monitoring technologies are widely used by consumers, clinicians, and researchers to track personal, patient, or test subject status. A critical part of assessing status is having the necessary tools to extract and interpret the veritable mountain of time series data. Commercial time series data management tools are generally expensive and awkward to use. This report describes the development of Data Viewer/Data Miner (DVDM) software tools designed to facilitate time series data management, particularly as it relates to field studies of soldiers and Marines under the Warfighter Physiologic Status Monitoring (WPSM) program.

The WPSM program is a multi-institute research program focused on developing a suite of wearable sensors to provide critical physiologic state information to commanders and medics. The sensor-hardware and network provide a wealth of physiologic information within a broader framework of clothing, weather, and mission. These data are used to identify which sensors provide critical and useful information, and to develop models of thermal stress, hydration status, metabolic requirements, and cognitive state. The WPSM program is being developed in an iterative manner, utilizing the test-model-test methodology (13, 14, 15). Each testing cycle can mean a large field study, with multiple WPSM systems garnering days of data from subjects.

DATA VOLUME PROBLEM

To researchers accustomed to utilizing simple office spread sheets or statistical packages to manipulate and analyze data, the sheer quantity of data generated from a WPSM system can be overwhelming. In a typical WPSM experiment (WPSM Winter 1999, Quantico, VA), the system generated over 216,000 individual records of data with at least 7 fields per record, for a total of 1,512,000 data points. As the number of subjects rise and study durations lengthen, these numbers will increase. Lewis (17) suggests that the bottleneck in scientific production is not in the basic science, but in the ability to deal effectively with the data. This is clearly true for the WPSM program.

BENEFITS OF AUTOMATION

The goal of automation is to move the bottleneck back to the basic science and away from data handling. This should mean that researchers have access to pertinent data quickly and effectively. Furthermore, time is not wasted fumbling with ineffectual tools; formatting and reformatting data; merging, splitting and sorting files; and continually rewriting software. The ability to deal with data in an automated and timely manner is critical to the WPSM development process. This paper details the development of an automated data management architecture and discusses its application to a number of data sets.

METHODS

A comprehensive literature review was conducted to understand how other researchers approached the problem of biologic data management. Appendix (A) contains details of the databases and key words used in the search. The review identified the following: a number of different information management fields; discussions over which archiving or database management tools are best to use; and debate about the efficacy of different database models. A common three-step information management theme emerged from the review: (1) Characterize, (2) Archive, and (3) Retrieve. Automation is only achieved when attention has been paid to these three areas. Basically, data need to flow automatically from their source to an archive to be accessed simply by an experimenter. Characterization is important to define both the architecture of the archive and the software used to connect the data sources to the archive. Archive considerations affect how easily data can be stored and retrieved. Retrieval software interfaces with the archive and provides the interface between a researcher and the underlying archive data structure. A poor interface can make it almost impossible to effectively query the data set.

DATA CHARACTERIZATION

For a typical WPSM field study, many different types of data can be collected. Each type of data can have its own collection interval, duration of data validity, and a person or thing to which the data relate. Table 1 shows typical data from a WPSM field study.

Table 1: Typical WPSM Field Study Data

<u>Data Type</u>	<u>Collection Interval</u>	<u>Data Persistence</u>	<u>Data Relate to</u>
Physiologic	1 min	Instant	Subject
Meteorologic	15 min	15 min	Met. Stations
Clothing log	60 min	60 min	Subject
Dietary	24 hour	24 hour	Subject
Equipment log	60 min	60 min	Subject
Video	Sporadic	Length of clip	None/Some/All
Photographic	Sporadic	Instant	None/Some/All
Activity log	60 min	60 min	Subject
Geo-Location	1 min - 2 sec	Instant	Subject
Weight	24 hour	24 hour	Subject
Biographic	Once	Study	Subject
Comments	Sporadic	Instant / Range	None/Some/All

Although these data are complex, they would not pose a problem for automation, since all the data are known. However, the variable nature of field studies makes these characterizations difficult. For example, the number and type of sensors can vary according to new developments and new areas of research. From study to study, characterization of data becomes an open-ended question, leading to a "chicken and egg" conundrum. That is, data need to be characterized to develop an effective data archive structure, with automated interfaces between the data and archive. However, without a solid characterization of the original data, developing an archive structure is difficult.

Our first challenge was to develop a comprehensive method of representing all study data in a standardized way. This method had to be extensible without demanding archive architectures and interface software to be rewritten every time new data were introduced.

Data Standardization

In order to help with the standardization process, object-oriented principals (7) were used. All study data types were abstracted and broken down into key components. Through this process, five identifying properties were found to characterize each data point: location, time, temporal persistence, to whom or what the data related, and what the data represented. These five properties allowed the data to be represented by three key axes on which data can be collapsed or expanded: space (geo-location), time, and entity. This object-oriented representation provided an extensible but characterizable means of representing WPSM field data.

The newly formed Extensible Markup Language (XML) (25) provided a way to implement the WPSM standardized data. This technology is based upon open flat files with semantically tagged data items. Tagging of data is not new and has been employed in the Standard Generalized Markup Language (SGML), an international standard (12, 25) widely used for document management. This technique has also been used with some success in certain bibliographic and complex protein sequence databases (1, 4, 9), along with USARIEM's MERCURY system, a weather-mission planning tool (10, 11, 19, 20).

Aside from being a good format to represent WPSM data, XML has been adopted by the World Wide Web Consortium (W3C, <http://www.w3c.org>, <http://www.w3c.org/XML/>) as a standard for data exchange. This has meant that an increasing number of companies are developing commercial off-the-shelf (COTS) XML tools, examples of which include parsing tools (JAVA <http://java.sun.com>), display tools (Internet Browsers), and editing tools (IBM Xeena <http://www.alphaworks.ibm.com/tech/xeena>). XML offers further advantages, as a number of efforts are underway to allow XML files to be searched or queried in a similar way to commercial databases: XQL (22), QUILT (23), XML-QL (8), Xpath (27), YATL (5).

Thus, the XML format seemed a timely and natural choice for an extensible and standardized way of representing our field study data streams. Figure 1 provides an example of one piece of XML data. Appendix B details more fully all the encoding of WPSM data, and Appendix C details the WPSM data dictionary for the U.S. Marine Corps Infantry Officer Course, Quantico, VA, field study.

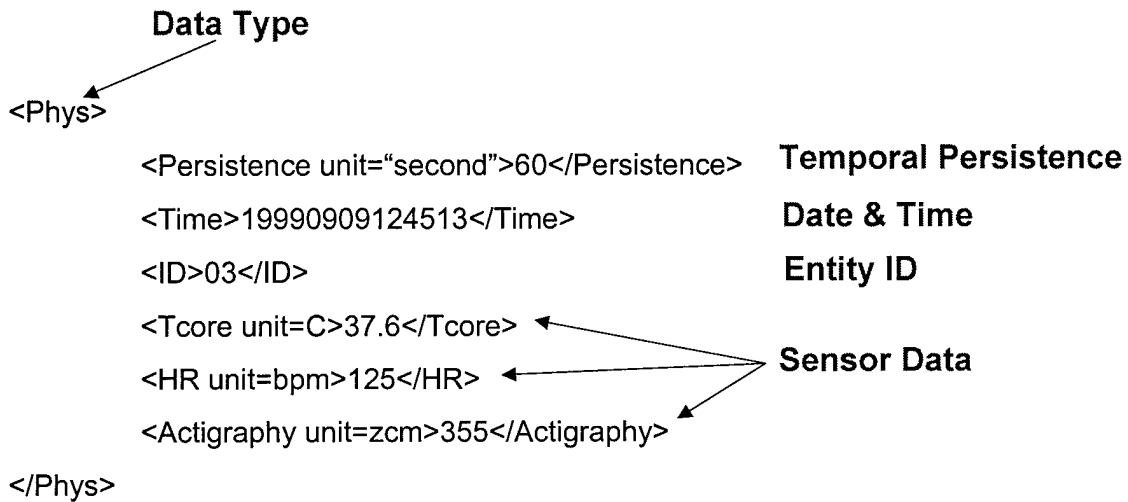


Figure 1: WPSM XML Data Example: This graphic shows one piece of physiologic data from a field study. The `<Phys>` tag identifies the data type. Within the data type, `<Time>`, `<Persistence>`, `<EntityType>`, and `<EntityId>` are required attributes. These identify the time at which the data were collected, how long the data remain valid, and to what device or person the data belong. The remaining tags identify measured parameters.

DATA ARCHIVE

Most data archiving schemes are based on commonly used models, such as relational (2, 3, 17, 18, 26), object/relational (2, 3, 17, 18, 26), flat file (2, 9, 17, 18, 26), hierarchical (26), and hybrid (2) schemes. Each model has its advantages and disadvantages for archiving data. Figure 2 attempts to provide data model selection criteria based upon data and query complexity.

The object methodology used to solve the data standardization problem fits very well with the object relational style data model. WPSM data are indeed complex, as are the types of questions that can be asked by the scientist. However, the object relational model does not provide the entire solution. The time series nature of most WPSM data, and the temporal nature of most queries, present difficult challenges for most database management systems (DBMS). While object relational databases are improving and time series manipulations can be done within database objects (Oracle, Redwood Shores, CA, <http://www.oracle.com>) and Informix (IBM Corporation, White Plains, NY, <http://www-4.ibm.com/software/data/informix/>), this solution is not ideal. These systems were not built with time series manipulations as a goal, and the underlying access

graphical user interface to allow scientific questioning of the data. Most functionality of the developed code would ultimately be necessary in an object-relational DBMS. Software to read and preprocess the WPSM XML data would be needed to provide a link to any database. The database itself would also need a graceful graphical user interface. Thus, any investment in custom code would be necessary, to some extent, for any archive or database scheme.

WPSM Data Archive

Base Directory

Where

When

XML Tagged Data

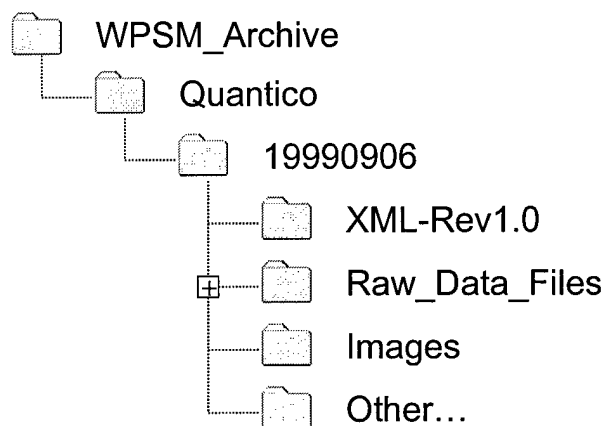


Figure 3: WPSM XML Sample Data Archive Structure

Having decided upon the flat file approach with XML data files, the archive structure itself was relatively simple. Figure 3 shows a basic archive structure for one experiment. The archive itself is based upon four directory levels. The first level is a directory for the archive or collection in which all applicable studies, experiments, or sets of data reside. The second directory level describes the location of where the data were collected. The third level details the time when the data were collected. The second and third directory levels could be swapped. The fourth directory level contains various directories for data. One directory contains data in the standardized XML format. This may be one XML file or multiple files, depending on experimenter choice (see Appendix B). Other directories contain raw data files, images, and comments. These files and images can be referenced from the XML data.

Context and Revision Control

Context. Since study data are being stored in the World Wide Web Consortium (W3C) standard for data exchange, a simple way to provide context for the experimental data is to utilize the ubiquitous hyper-text markup language (HTML) standard for documents (<http://www.W3C.org>).

Simply, an experiment can be described in the usual scientific manner, and the technical report or journal article converted to an HTML document (MS Word and Word Perfect can automatically convert files to the HTML format). These HTML documents could then be placed in the directory for that experiment and be read by any web browser.

Revision Control. Even with an automated data collection system, data smoothing, cropping of outliers, and the addition of derived values will still be needed. In the directory system, this is accomplished by creating a new data directory for each new revision. Within the directory, an HTML document will detail what changes were made to get to the current version of the data. This method will ensure all versions of data remain intact and that there is a clear and documented path from the raw data to the most current revision.

DATA ACCESS

As detailed in the previous two sections, primary data access is provided through a custom application. The choice of XML as the standard file format also allows the use of other XML data viewing and data access tools.

To enable the most extensibility and code usefulness, the data access code was written using a client-server architecture. Figure 4 shows a block functional diagram of the architecture.

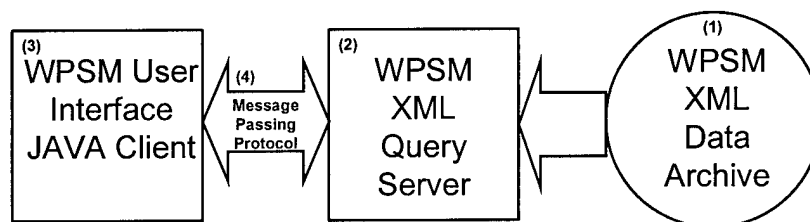


Figure 4: Block Functional Diagram of Data Access Application Architecture. (1) The WPSM XML data archive can be any file or group of files within a single directory that conform to the WPSM XML standard. (2) The WPSM XML query server understands the WPSM XML file format, accepts queries via a message passing protocol, and returns query results. (3) The WPSM user interface, written in JAVA to enable web browsing, allows queries to be generated in an intuitive and simple manner. (4) A standardized message passing protocol allows any program to request data from an XML archive.

There are three main components to the entire archive: a WPSM XML data archive, a WPSM XML query server, and a WPSM user interface. Each of the three components can reside on a different computer and be accessed through a network. In theory, this mode of operation can be extended to the Internet; however, there can be scaling problems associated with very large query results being passed across servers.

WPSM XML Query Server

The XML query server provides a generic interface to the XML data archive. The server has several functions. The first function is to enumerate a WPSM XML archive. This provides basic archive information, such as what data types and data elements are present, and how many individual subjects exist. This type of information is utilized by the WPSM user interface to provide context for queries. Once an archive has been enumerated, the server can accept queries and output results based upon the contents of the archive. The server is also designed to specifically deal with time series and temporal style data. This means that data on different time lines and different time resolutions can be queried together; missing data points can be interpolated by a number of methods; and subject information can be collapsed across time.

The query server will also function independently of the graphical user interface. The server will accept queries and output results through simple ASCII text files. Thus the server can be utilized by other programs or even accessed directly by using the message passing protocol and a text editor. (See Appendix E for a more thorough presentation of the query message passing protocol.)

WPSM User Interface Client

The WPSM user interface client was built using JAVA technology to enable the application to run on multiple platforms and provide the possibility to run within web browsing software. The interface allows a researcher to open any WPSM XML archive, view the range of data types and data elements, and graphically generate queries to extract data. After data have been queried, the results are tabularized and, if possible, plotted. Once the user has identified a pertinent period or sample of data, the results may be exported into a comma separated value (CSV) file. This file format is readily importable into common spreadsheet packages and other analysis tools. The user interface also provides a trail of queries allowing the user to note a profitable path of data mining or research.

The goal of the user interface is to allow a researcher to quickly view and select data from a number of different sources, find interesting periods of data, and export these data to other tools for further analysis. The user interface also provides the mechanisms to help organize and filter the data sets.

Figure 5 shows the basic user interface screen, where the subject can ask for data based on a number of selection criteria.

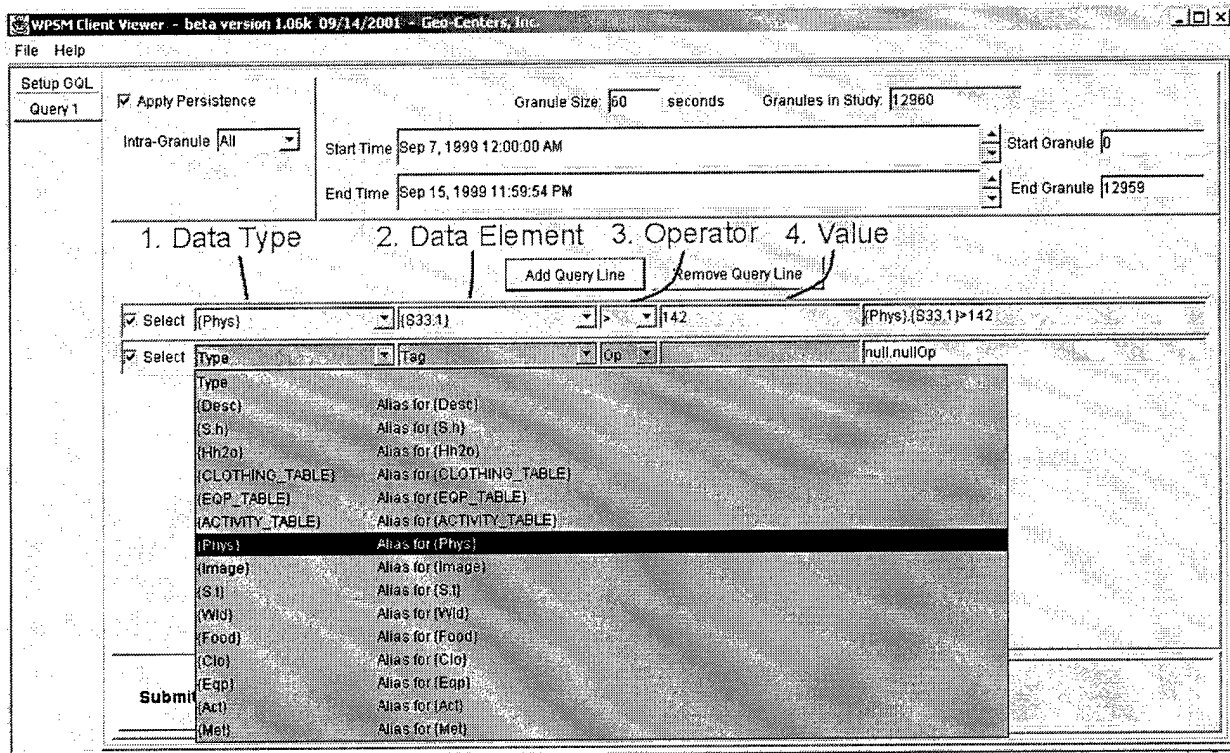


Figure 5: User interface showing data type selection.

To generate a query, users must first select the type of data they wish to query (Figure 5, #1.). Once a data type has been selected, a data element, such as heart rate, is chosen (Figure 5, #2.). Next, a data operator should be selected along with an appropriate value (Figure 5, #3-4). Supported operators include <, <=, =, NOT =, >, >=, Between, NOT Between, and All. Subsequent query lines are logically ANDed together. The query in Figure 5 requests all data for type {Phys} and element {S33.1} (a heart rate sensor unit), where values are greater than 142. Figure 6 shows a graphical presentation of a typical query result.

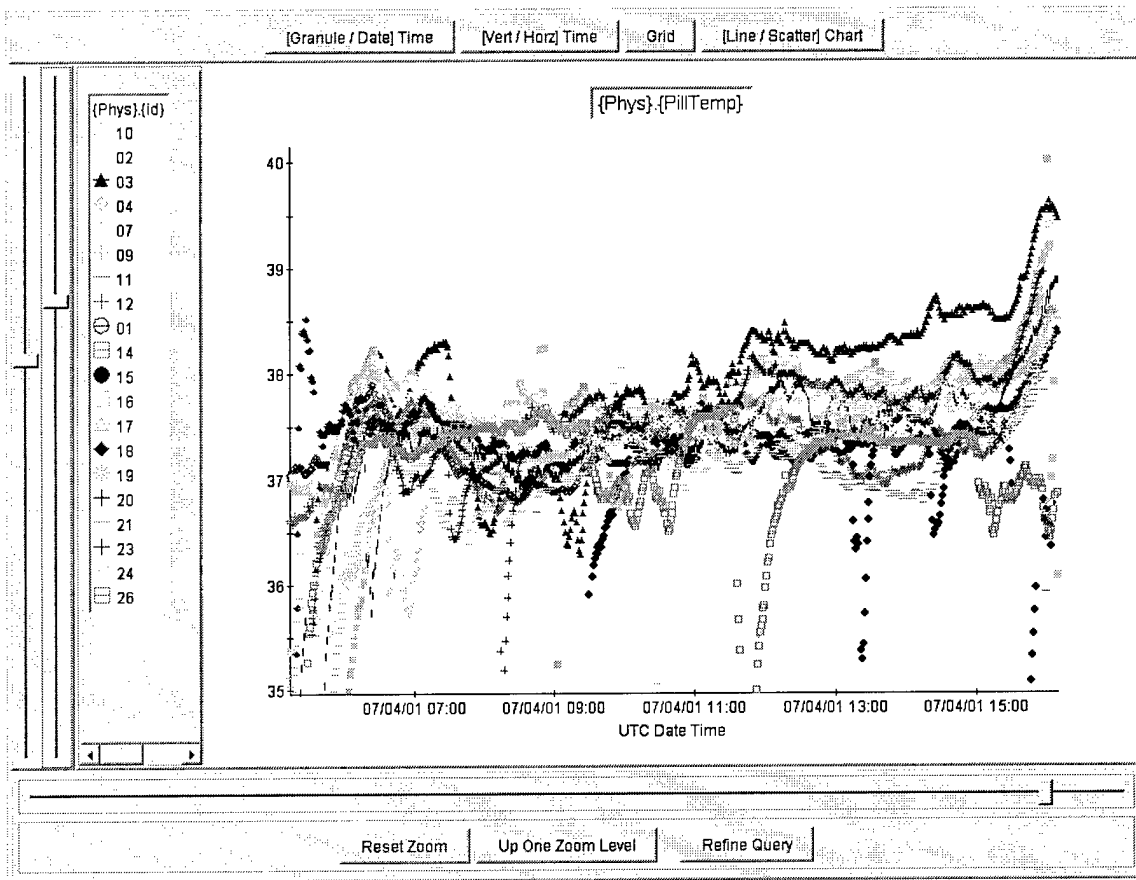


Figure 6: Graphical Display of Query Results

RESULTS

An alpha version of the custom software called Data Viewer Data Miner (DVDM) became available during spring 2001 (GEO-CENTERS INC., 7 Wells Avenue, Newton, MA). The first set of data to be archived utilizing the standardized WPSM XML was the September 1999 U.S. Marine Corps Infantry Officer Course data. These combined data created the largest archive for WPSM to date, and provided a test data set for the new software. In XML, the data set totaled 33 Mb, excluding picture and video files.

U.S. MARINE CORPS INFANTRY OFFICER COURSE (SEPTEMBER 1999)

This first encapsulation of field study data was very revealing, as the DVDM opened a door to data visualization not possible before. In the past, multiple charts of data were produced for each subject, for each parameter, for each day. Comparing data amongst subjects and days was time-consuming and difficult. DVDM enabled data to be displayed simultaneously for selected parameters and subjects. Scientifically interesting periods of physiologic response could be requested quickly. Questions such as, "Find me all periods where the heart rate is above 140," became easy to ask. Once data from these types of requests were returned, the researcher could zoom in on various time periods, and further investigate what other parameters or events had an

impact on or were present during this period. DVDM provided rapid access to a large data set. The program's ability to time-correlate data and subjects, and export this data, enabled the researcher to concentrate on analysis rather than on data access and formatting.

U.S. MARINE CORPS INFANTRY OFFICER COURSE (JULY 2001)

The successful implementation and use of the first data set in XML demonstrated the usefulness of the approach as a whole. To facilitate capturing data in XML, a parsing program was written to allow the conversion of all current WPSM raw data formats to WPSM XML. This software was utilized for a field study with the U.S. Marine Corps Infantry Officer Recruits during July 2001. The study examined 27 subjects over a five-day marksmanship-training course. The parsing software was applied to all field study sensors as they were downloaded, enabling a comprehensive data set to be assembled as soon as data were available. Researchers were thus able to quickly identify problems with data and sensors, minimizing data loss. It also allowed for a report to be compiled and presented by the end of the study.

U.S. CENTER FOR ENVIRONMENTAL HEALTH RESEARCH (USACEHR)

USACEHR has data management problems similar to the WPSM program. USACEHR measure physiologic signals from fish and collect water quality information. The generic ability of the DVDM to deal with new and different types of data was proved by applying the XML technology to USACEHR's data. DVDM was delivered with a package of tools to enable XML storage of both real time fish data and model outputs. The package also provided parsing routines to convert and archive old data. DVDM provided a useful tool to examine new and old data sets. The software included a simple model of fish strain. Model parameters could be altered and results entered into the XML archive. New and old data sets could be run through the model and the results viewed simultaneously in DVDM. The ability to compare model outputs with actual data will help increase the frequency of the test-model-test cycle.

SCENARIO-J

The open architecture of the XML archive and DVDM provided opportunity to automate the link from field study data to models. The SCENARIO model provides simulations that generate the time course of body temperature shifts, thermoeffector responses, and central and peripheral circulatory changes (17). The model can also act as a prediction tool, estimating physiologic responses based upon predicted environment, workload, and clothing ensembles. Figure 7 shows an architecture providing a link from an XML archive to a JAVA implementation of the SCENARIO model, SCENARIO-J.

The three original WPSM DVDM and data archive components are still present (Figure 7, #2, #3, and #4). Middleware software and the SCENARIO model were added (Figure 7, #1, #6). The middleware software utilizes the message passing protocol and

the XML query server directly without the WPSM user interface (Figure 7, #5). The SCENARIO model user interface was altered to allow the user to select the desired XML data archive. The model requests data from the middleware software, which generates a series of queries for the XML query server. Data are passed back to the middleware software and run in an iterative manner through the model. Output from the model is inserted into the XML archive.

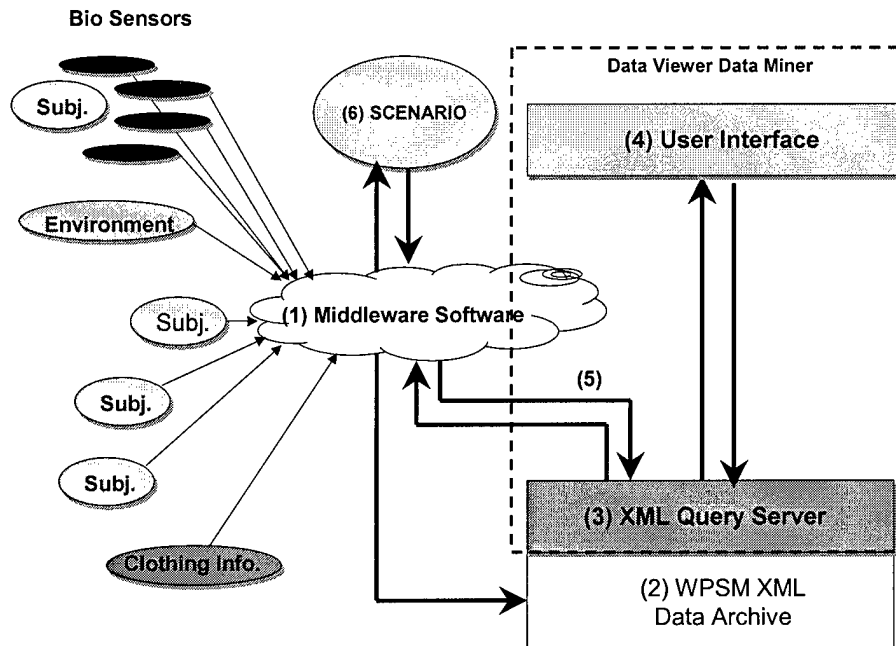


Figure 7: Block functional diagram of WPSM data management architecture with link to SCENARIO Model

Data from the U.S. Marine Corps Infantry Officer Course (September 1999) has been successfully run through this model. Model output and actual data input can be viewed simultaneously with DVDM.

DISCUSSION

The WPSM data automation work has developed a number of useful strategies and components. A standard method for describing and representing field study data through XML has been developed and implemented, solving the main problem of automation and extensibility. With a common format for field study data, a software suite was developed to solve time series and temporally based data problems. The software was also designed to provide a rapid means of viewing, querying, and manipulating data. The tools are generic by design, and as long as data are represented in WPSM encoded XML, they will work with any new data sets. The standard XML data format has been applied to three WPSM data sets. The DVDM tool has provided rapid access to data, and increased the speed of formatting data for analysis. The whole data management strategy and architecture has proven useful to other researchers, such as those at USACEHR. There the researchers have different

experimental designs to WPSM, yet the system proved powerful enough to improve their data management. The basic architecture has allowed an automated link from field study data to models and their outputs, speeding the test-model-test cycle.

LIMITATIONS

Although this architecture has proved useful, there are basic limitations. XML is a verbose file format, so where data are collected with high sampling rates, large files are produced. This could pose a problem, depending on computer capability. The DVDM tools resolve temporal issues by loading the entire selected data archive into memory. So the larger the data archive, the more memory is needed. For the WPSM field studies, a capable computer (e.g., Intel Pentium 3, with 450 MHz clock speed and 356 Mbytes of RAM) is sufficient for the ~1.5 million data points which were generated. Although existing WPSM data is acquired with a relatively low sampling rate (minutes not seconds), data at a higher sampling rates can be viewed. However, there is a trade off between sampling rate and the duration of a study. Increasing the sampling rate decreases the total amount of time DVDM can handle in a set amount of memory. Increasing memory or dividing studies into smaller chunks of time can alleviate this problem.

ETHICAL USE OF ONLINE DATA

As more and more data sets become digitized and made accessible to a broader audience, consideration should be given to the ethics of public access and use. Davis et al. (6) provide a suggested code of ethics to apply to the disposition of all on-line raw data sets.

CONCLUSIONS

The standardized field study XML data format and the DVDM tools have proven useful to the WPSM program and to other users. They have allowed the automation of data collection and archiving, and have enabled the rapid viewing and analysis of data sets. For WPSM, this automation strategy has moved the bottleneck back from data management to the basic science.

REFERENCES

1. Baclawski, K., R. Futrelle, N. Fridman, and M. J. Pescitelli. Database techniques for biological materials and methods. *ISMB* 1: 21-28, 1993.
2. Blake, J., and C. J. Bult. Biological databases on the Internet. In: *Molecular Zoology: Advances, Strategies, and Protocols*, edited by J. D. Ferraris and S. R. Palumbi. Wiley-Liss, Inc., 1996.
3. Combi, C., and Y. Shahar. Temporal reasoning and temporal data maintenance in medicine: issues and challenges. *Comput. Biol. Med.* 27(5): 353-368, 1997.
4. Coupaye, T. Wrapping SRS with CORBA: from textual data to distributed objects. *Bioinformatics* 15(4): 333-338, 1999.
5. Cluet, S., S. Jacquemin, and J. Simeon. *The New YATL: Design and Specifications*. INRIA. Technical Report, 1999.
6. Davis, M. A., D. Tilman, S. E. Hobbie, C. L. Lehman, P. B. Reich, J. M. H. Knops, S. Naeem, M. E. Ritchie, and D. A. Wedin. Public access and use of electronically archived data: ethical considerations. *Bulletin of the Ecological Society of America*: January 2001.
7. Deitel, H. M., and P. J. Deitel. *C++ How to Program*. Englewood Cliffs, NJ: Prentice Hall, 1994.
8. Deutsch, A., M. Fernandez, D. Florescu, A. Levy, and D. Suciu. A Query Language for XML. See <http://www.research.att.com/~mff/files/final.html>.
9. Etold, T., and P. Argos. SRS – An indexing and retrieval tool for flat file data libraries. *Comput. Appl. Biosci.* 9(1): 49-57, 1993.
10. Fields, C. A., J. E. Newberry, H. D. Pfeifer, C. A. Soderland, S. F. Kirby, and G. B. McWilliams. MERCURY: A heterogeneous system for spatial extrapolation of mesoscale meteorological data. *International Journal of Man-Machine Studies*, 36: 309-326, 1992.
11. Furlong, J. R. Integration of Warfighter Physiologic Status Data with a Battle-Scale Weather and Terrain Information System - Final Report. Frederick, MD: SAIC. Technical Report DAMD 17-98-D-0022, 1998.
12. Goldfarb, C. F., and P. Prescod. *The XML Handbook (2nd Ed)*. Upper Saddle River, NJ: Prentice Hall, 2000.

13. Hoyt, R. W., M. J. Buller, M. S. Redin, R. D. Poor, S. R. Oliver, W. T. Matthew, W. A. Latzka, A. J. Young, D. Redmond, and C. Kearns. Soldier Physiological Monitoring - Results of Dismounted Battlespace Battle Lab Concept Experimentation Program Field Study. Natick, MA: USARIEM. Technical Report T98-6, November 1997.
14. Hoyt, R. W., M. J. Buller, J. P. DeLany, D. Stultz, K. Warren, M. P. Hamlet, D. Shantz, W. T. Matthew, W. J. Tharion, P. Smith, and B. Smith. Warfighter Physiological Status Monitoring (WPSM): Energy Balance and Thermal Status during a 10-day Cold Weather U.S. Marine Corps Infantry Officer Course Field Exercise. Natick, MA: USARIEM. Technical Report T-02/2, October 2001.
15. Hoyt, R. W., M. J. Buller, M. P. Hamlet, W. T. Matthew, W. J. Tharion, and J. P. DeLany. Thermal Stress Indices in U.S. Marine Corps Infantrymen during a Cold Weather Field Exercise (FEX). *FASEB J.* 453(3): A612, 2000.
16. Kraning II, K. K., and R. R. Gonzalez. A mechanistic computer simulation of human work in heat that accounts for physical and physiologic effects of clothing, aerobic fitness, and progressive dehydration. *J. Therm. Biol.* 22(4/5): 331-342, 1997.
17. Lewis, S. Design issues in developing laboratory information management systems. In: *Automated DNA Sequencing and Analysis*. Academic Press, 1994, pp. 329-338
18. Lindek, S., R. Fritsch, J. Machtynger, P. A. de Alarcón, Chagoyen. Design and realization of an on-line database for multidimensional microscopic images of biological specimens. *J. Structural Biol.* 125: 103-111, 1999.
19. Matthew, W. T., W. R. Santee, R. W. Hoyt, P. Tikuisis, E. S. Barnes, G. B. McWilliams, H. D. Pfeiffer, and J. Furlong. Automated thermal injury risk assessment in the dismounted infantry battlespace. In: *Proceedings of the 1996 Battlespace Atmospheric Conference*. San Diego, CA: NRAD, 1996, p. 579-586.
20. Matthew, W. T., R. W. Hoyt, T. J. Doherty, M. J. Buller. Integration of weather effects models with real-time physiological measurements in the dismounted infantry battlespace. In: *Proceedings of the 1997 Battlespace Atmospheric Conference, 2-4 December 1997*, edited by K. D. Anderson, and J. H. Richter. San Diego, CA: Space and Naval Warfare Systems Center, 1998, p. 653-660.
21. Melton, J., and A. R. Simon. *Understanding the New SQL: A Complete Guide*. San Francisco: Morgan Kaufmann Publishers, 1993.
22. Robie, J., J. Lapp, and D. Schach. *XML Query Language (XQL)*. <http://www.w3.org/TandS/QL/QL98/pp/xql.html>.
23. Robie, J., D. Chamberlin, and D. Florescu. *Quilt: an XML query language*. <http://www.gca.org/papers/xml europe2000/papers/s08-01.html>.

24. Snodgrass, R. T. *Developing Time-Oriented Database Applications in SQL*. San Francisco: Morgan Kaufmann, 2000.
25. St. Laurent, S. *XML Primer* (2nd Ed.). New York: Hungry Minds Inc., 1999.
26. Winslett, M. New database technology for nontraditional applications. In: *Advances in Computer Methods for Systematic Biology: Artificial Intelligence, Databases, Computer Vision*. Baltimore, Maryland: Johns Hopkins University Press, 1993.
27. Xpath World Wide Web Consortium. *XML Path Language (Xpath) Version 1.0*. W3C Recommendation. <http://www.w3.org/TR/xpath.html>.

APPENDIX A LITERATURE REVIEW

METHODS

A literature review was conducted to assess how other researchers approach the problem of biological data management. The search was designed to be broad initially, to ensure that most pertinent areas of research could be identified, and then honed down to specific areas of interest. The search was conducted in an iterative manner. Initial search results were used to order likely articles, and these helped identify other papers and other areas of research.

Information was derived from scientific journal articles, government and non-government reports, databases, Internet sites, other project-related manuscripts, and correspondence with primary investigators.

All Literature search results were reviewed, and any applicable articles were ordered and studied. All reviewed references were broken out into the following categories: BioInformatics/Computational Biology, Temporal Databases, Software Design/Programming Languages, Information Broker Architecture, and Medical Informatics.

Database Searches and Keywords

Internet Grateful Med V2.6.3 (<http://www.ncbi.nlm.nih.gov/entrez/>) was searched using combinations of the following key words:

- Biological Databases
- BioInformatics
- Data Mining Approaches

MEDLINE database (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi>) was queried using the following key terms:

- Computational Biology
- Data Integration
- Database Management Systems

The FirstSearch Catalog (FS) database. (<http://www.firstsearch.com>)
Keywords for the queries were combinations of the following groupings (some searches were limited to the years 1997–present):

Computer Programming Language

Relational Databases
Management
Standards

Object-Oriented Programming Computer Science
Standards

Programming Languages Electronic Computers
Standards

Information Technology Industry Council

SQL Computer Program Language

Temporal Databases

Computational Biology

Information Brokerage Architecture
Data Brokerage
Data Architecture

PubMed Central (<http://www.pubmedcentral.nih.gov/>), part of NCBI; to access free full-text articles.

Web-based Patient Data
Clinical Data Retrieval
Decision Support Systems

The National Center for Supercomputing Applications' (NCSA) database
was queried using the following keyword combination:

Biological
Database Management

These initial searches revealed many more pertinent keywords, and helped us expand our search. Below is a comprehensive keyword listing. Once papers were ordered and reviewed, other articles were identified from reference lists and bibliographies. The pertinent results of the literature review can be found in the bibliography section of this report.

List of Keywords Used in Literature Search

Architecture
Bioinformatics
Biological
Biological Databases
Clinical Data Retrieval
Component Based Medical Decision Support System
Computational Biology
Computer Programming Language
Computer Communication Networks
 Computer Graphics
 Decision Support Techniques
 Human
 Patient Care Planning
 Patient Education
 Remote Consultation/Instrumentation
 Therapy, Computer-Assisted
 User-Computer Interface
Data Agents
Data Architecture
Data Brokerage
Data Compression
Database Design
Data Dissemination
Databases, Factual
Data Granularity
Data Integration
Data Mining
Data Mining Approaches
Data Mining Tools
Data Warehousing
Database Management Systems
Decision Support System
Disaster Medicine
Dynamic Linking
Enterprise Information Networks
Graphical User Interface
Image Processing
Imaging System
Information Agents
Information Brokerage Architecture
Information Storage and Retrieval
Information Technology Industry Council
Internet

Internet Applications
Knowledge-based Systems
Medical Informatics
Medical Monitoring
Middleware
Mobile Computing
Object Framework
Object-Oriented Programming Computer Science
Palm-top Digital Assistant (PDA)
Programming Languages Electronic Computers
Query Language
Relational Databases
Sensors
SQL Computer Program Language
Standards
Telecommunications
Telemedicine
Temporal Abstraction Knowledge
Temporal Databases
Temporal Data Models
Time-Constraints Databases
Transaction Time
User-Defined Time
Valid Time
Visualization
Web-based Patient Data
Wireless Networks, Self-Organizing

Projects and Products

Internet search engines: Altavista <http://www.altavista.com>, Infoseek <http://www.infoseek.com>, Lycos <http://www.lycos.com>, and Google <http://www.google.com> were utilized to identify areas of research and COTS information management tools. The following is a list of available tools that were reviewed.

DXtractor (<http://www.chip.org/chip>) is an application that allows easy querying of a clinical database by clinicians; identifies patient populations; queries temporally; uses clinical abstractions; requires minimal computer expertise. See the following:

Nigrin, D. J., and K. Kohane. Temporal expressiveness in querying a time-stamp based clinical database. *J. Am. Med. Inform. Assoc.* 7: 152-163, 2000.

BLAST[®] (Basic Local Alignment Search Tool) is a set of similar search programs designed to explore all of the available sequence databases regardless of whether the query is protein or DNA. The BLAST programs have been designed for speed, with a minimal sacrifice of sensitivity to distant sequence relationships. See the following:

NCBI Tools for Bioinformatics Research. <http://www.ncbi.nlm.nih.gov/BLAST.html>.

The Kleisli System (<http://www.kris-inc.com/>) is a tool for complex queries across multiple databases and for data integration in biology. See the following:

Chung, S.Y. and L. Wong. Kleisli: a new tool for data integration in biology. *Trends Biotechnol Sep*;17(9): 351-5, 1999.

DBIS – Toolkit. The Dissemination Based Information System (DBIS) prototype toolkit is a set of modules which can be used to build-up a heterogeneous distributed client-server network, supporting different modes of data transfer (i.e., unicast, multicast, broadcast, push, pull); middleware for large scale data delivery. See the following:

Altinel, M., D. Aksoy, T. Baby, M. Franklin, W. Shapiro, and S. Zdonik. DBIS-Toolkit: Adaptable Middleware for Large Scale Data Delivery, Demo Description for ACM SIGMOD Conference, city?, PA, 1999. (<http://www.cs.umd.edu/~altinel/sigmod99/sigmod99>).

GENPRO provides automatic generation of Prolog clause files for knowledge-based systems in the biomedical sciences (e.g., protein structure prediction and modeling). See the following:

Saldanha, J., and J.R. Eccles. GENPRO: automatic generation of Prolog clause files for knowledge-based systems in the biomedical sciences. *Comput Methods Programs Biomed.* Mar;28(3): 207-14, 1989.

RESUME is a system that performs temporal abstraction of time-stamped data. The temporal-abstraction task is crucial for planning treatment, for executing treatment plans, for identifying clinical problems, and for revising treatment plans; generates temporal abstractions, given time stamped data and events. See the following:

Stanford University, Medical Informatics. <http://smi-web.stanford.edu/pubs/>.

Protégé is a general framework and set of tools for the construction of knowledge-based systems. See the following:

Shahar, Y., H. Chen, D.P. Stites, L. Basso, et al. Semi-automated Entry of Clinical Temporal-abstraction Knowledge. *Journal of the American Medical Informatics Association* 6(6): 494-511, 1999.

Research Institutes, Associations and Societies

This section provides information on a number of centers of excellence, which were identified in the literature review. Often the web sites of these organizations detail efforts in biologic information management.

American Medical Informatics Association, <http://www.amia.org>.

American Society for Information Science, <http://www.asis.org>.

Boston University, <http://bioinformatics.bu.edu/>.

Brown University, <http://www.cs.brown.edu/>.

POC: Stanley B. Zdonik, Jr.
Stanley_Zdonik_Jr@Brown.EDU
Computer Science Department, Professor
401-863-7648
Box 1910, Brown University, Providence, RI 02912-1910 US

Children's Hospital Informatics Program, <http://www.chip.org/chip/htmlindex.html>.

European BioInformatic Institute

Journal of the American Medical Informatics Association (JAMIA), <http://www.jamia.org>.

Mass. Institute of Technology, Clinical Decision Making Group,
<http://www.medg.lcs.mit.edu/>.

POC: Dr. Jon Doyle, doyle@mit.edu.

National Center for Biotechnology Information (NCBI), National Institute of Health (NIH), National Library of Medicine (NLM), <http://www.ncbi.nlm.nih.gov/>.

NCBI Tools for Bioinformatics Research, <http://www.ncbi.nlm.nih.gov/Tools/index.html>.

NCBI's PubMed Central, <http://www.pubmedcentral.nih.gov/>.

NLM: IGM Metathesaurus Information Screen, <http://130.14.32.42/cgi-bin>.

National Center for Supercomputing Applications (NCSA), <http://www.ncsa.>

Networked Social Science Tools and Resources (NESSTAR), <http://www.nesstar.org>.

Stanford University, Medical Informatics, <http://smi-web.stanford.edu/pubs/>.

POC: Mark Musen, musen@smi.stanford.edu.

APPENDIX B WPSM XML ENCODED DATA

BACKGROUND

XML, like Hyper Text Markup Language (HTML), is a subset of Standard Generalized Markup Language (SGML), the international standard for structured information (12). The WPSM program has utilized XML to encode field study data in an extensible way. This appendix will detail the structure and encoding of WPSM data within XML, but will provide little detail of XML itself. For further XML information, see references 12 and 26.

SOME BASIC XML

XML uses tags to identify data within a file. Any piece of data must be enclosed between start and end tags, such as the following:

```
<something>data</something>
```

The tags can be anything that describe what the data element is, such as the following:

```
<corebodytemperature>36.7</corebodytemperature>
```

Data elements can also be nested, for example:

```
<somedata>  
    <subject>01</subject>  
    <corebodytemperature>36.7</corebodytemperature>  
    <some_other_measurement>data</ some_other_measurement>  
</somedata>
```

In the previous example, the data element <somedata> contains three other data elements. In XML the element that has nested data within it is called a "node." Each piece of data within a node is called an "element." WPSM encoded data make use of nesting to generate data objects. All WPSM data are encoded as data objects. The data object definition is explained below.

For a file to be valid XML, the first line of the file should define it as such. The standard method is to use the following:

```
<?xml version="1.0" encoding="UTF-8"?>
```

All data within an XML file also need to be enclosed within a node. Thus, for the WPSM XML, the following node definition was chosen:

```

<STFF-XML version=1.0>
  ...
  (All WPSM XML data objects are in this node)
  ...
</STFF-XML>

```

Figure B1 shows a small but complete WPSM XML file.

Figure B1: Complete WPSM XML Data File

```

<?xml version="1.0" encoding="UTF-8"?>
<STFF-XML version=1.0>
<Phys>
  <Time>19990908050002</Time>
  <Subject>12</Subject>
  <Actigraphy units=zcm>123</ Actigraphy >
  <CoreBodyTemperature units=c>37.62</ CoreBodyTemperature >
</Phys>
</STFF-XML>

```

DATA OBJECT DEFINITION

WPSM data objects have one level of nesting or one node. The first level defines the data type of the object, while the elements within the node provide critical object information plus any other number of data elements. A fully defined WPSM data object would take the form shown in Figure B2.

Figure B2: Fully Defined Data Object

<u>Node</u>	<u>Elements</u>
<ObjectDataType>	<Persistence>length_of_time_in_seconds</Persistence> <Time unit=GMT localOffSet=-5:00>YYYYMMDDhhmmss.nnnnnnn</Time> <ID>data</ID> <LAT unit=degrees></LAT> <LON unit=degrees></LON> <ALT unit=meters></ALT> <userdefinedTag1 unit=something>data</userdefinedTag1> ... <userdefinedTagN unit=something>data</userdefinedTagN>
</ObjectDataType>	

Object Data Type

The node of any data object defines the type of data contained within the object. Figure B2 shows the node as <objectType>. This tag is user definable, and should have a valid XML tag to best define the type of data contained within the object. Thus, if physiologic data were being collected, the tag name could be <physiologic>. Where many different types of data are being collected, it is best to distinguish amongst these by type, collection rate, and to whom or what the data belong.

When coding data, it is important to properly distinguish between data types. There are several components to consider when determining a distinct data type.

Intuitive Component. In many cases, data types will be fairly obvious, (e.g., physiologic and meteorologic). Data types will be the large classes of data that may be examined independently.

Collection Interval. The collection interval becomes important in closely defining data types. For example, subject biographical data may be collected throughout a study. For WPSM, semi-nude weights, loaded weights, and circumferences were collected. Circumferences and semi-nude weights were collected at the beginning and the end of the study, while loaded weights were collected every day. Although these data at first seem to be of the same type, they should be split into different types, as their collection interval differs.

Data Persistence. The persistence of data is also a data type distinguisher. Again, using the biographic information as an example, the semi-nude weights and circumferences have a persistence of about 24 hours, while the loaded weight has a "sticky" persistence (i.e., the loaded weight is valid until it changes).

Entity Relationships. The final consideration of a data type is to whom or what the data in an XML object type belong. In the WPSM data, meteorologic data are collected on data loggers from weather stations at specific locations. Thus, the data relate to a particular weather station at a particular location. Physiologic data are collected on a WPSM system which is assigned to a warfighter. The physiologic information is related to a data logger identifier, rather than a subject. Other information within the field study was related directly to subject.

Persistence

Persistence is an optional object data element, and defines the amount of time that data within the object are valid. When persistence is not included, it is assumed that the data exist only for an instant in time. Persistence values are defined in seconds, with as much precision as necessary. An example of where persistence can be useful is where data are collected at different intervals but need to be accessed together. For example, weather data may be collected every hour, and physiologic data may be collected every minute. The weather data can be said to be valid for 59 minutes after it

is collected. Thus, the weather data collected at 12:00 is still valid at 12:59. It has a persistence of 59 minutes or 3540 seconds.

There are two special cases of persistence:

@STUDY
@STICKY

@STUDY denotes data that are valid from the point at which they were collected to the end of the study.

@STICKY denotes where data are valid until a new reading is taken. For example, in the WPSM study, weights were taken every day, but not necessarily at the same time of day. Hence, a weight for each subject is valid until a new weight for that subject is taken.

Time

Tag: Time

Attributes: units, localOffset, certainty

The date-time stamp records the point in time when the data within the object are valid. This may be the point when the data are recorded, but this may not necessarily be the case. Date and time are recorded in the following format:

YYYYMMddhhmmss.s

YYYY = year
MM = month
dd = day
hh = hour (24 Hour Notation)
mm = minutes
ss.s = seconds

Seconds can be of variable precision. This may mean some records may only have precision to the whole second (e.g., yyyyMMddhhmmss); or some may provide precision to many decimal places (e.g., yyyyMMddhhmmss.sssssssssssssss). It is necessary for a valid date time stamp to represent all times to at least the level of whole seconds. Date-time stamps must have at least 14 characters: four characters for the year, two for the month, two for the day, two for the hour, two for the minute, and two for the seconds. For example, if data on a line of STFF were collected at 1:45 pm on June 4, 2000, the date time field would be "20000604134500." Note that seconds are represented by "00."

Attributes. All the attributes for this element are optional. However, it is strongly advised that the *unit* and *localOffset* attributes are used.

Unit: Specify the time zone that is being used to represent the time data, such as Greenwich Mean Time (GMT) or Eastern Standard Time (EST).

LocalOffset: Note the actual offset between the recorded time and the local time using the *localOffset* attribute.

Certainty: This attribute allows a plus or minus time range in seconds to provide for measurements that have a period of uncertainty.

ID

Tag: ID

Attributes: - None -

The ID element identifies to whom, what, or what grouping the data in the object belong. The data of this element can be alphanumeric. Two reserved words can be used for the data in this element.

@NONE: Data are not attributable to anything.

@ALL: Data are attributable to all entities within the object.

Spatial Elements

Tag: LAT, LON, ALT

Attributes: - None -

To allow data objects to maintain a spatial component, the following three data element tags are used:

LAT: Latitude (degrees) North positive, South negative.

LON: Longitude (degrees) East positive, West negative.

ALT: Altitude/Elevation (meters above sea level).

User Defined Elements

Tag: - User Defined -

Attribute: unit

A data object can contain as many user defined data elements as needed. All user defined data elements have the same object attributes of time, persistence, ID, and location.

A user defined element can have any valid XML tag name except for WPSM reserved tag names. Data for each element can be both numeric and alphanumeric. Lists of data items can be generated, by using the same element name multiple times.

Units for each element should be defined in the *unit* attribute.

ASSIGNMENT OF DATA TYPES

Data types within WPSM XML files are independent. This means that ID element values within one data object type are not linked to ID element values within another. Assignment offers the mechanism for linking ID values from one data type to another. Assignments are accomplished using special data type data objects. The node tag <Assignment> is reserved as an assignment data type. Assignments can persist as any other data types, and have a specific time when they take effect. Assignments have no real ID element; however, one should be provided to allow assignments to be searchable data objects. Multiple and overlapping assignments are allowable.

Two types of assignments are provided: implicit and explicit.

Implicit Assignment

An implicit assignment links two data type entities directly. For example, WPSM data for clothing and activity are recorded by two different data types <Clo> and <Act>, respectively. Both clothing and activity data are related to each subject. Thus, the <ID> element in the <Clo> data type is the same as in the <Act> data type, (i.e., <Clo>.<ID>=<Act>.<ID>). Figure B3 shows an example of an implicit assignment.

Figure B3: Implicit Assignment

```
<Assignment>
  <Time>19990907000000.0</Time>
  <AssignmentMode>Implicit</AssignmentMode>
  <AssignedInputType>Subject</AssignedInputType>
  <AssignedInputTag>ID</AssignedInputTag>
  <AssignedOutputType>Weight</AssignedOutputType>
  <AssignedOutputTag>ID</AssignedOutputTag>
</Assignment>
```

Explicit Assignment

An explicit assignment links two different data types where the ID values may differ. For example, in the WPSM field studies, physiologic data are collected using on-body storage devices. These devices have their own ID numbers apart from subject ID values. Data collected from the storage devices are stored relating information to the

device and not the subject. However, the data from the devices relates to certain subjects at certain times. An explicit assignment maps the storage device ID element value to the subject ID element value. For example, <CollectionDevice>. <ID>30 maps to <Subject>. <ID>04. Figure B4 shows an example of an explicit assignment.

Figure B4: Explicit Assignment

```
<Assignment>
  <Time>19990907000000.0</Time>
  <AssignmentMode>Explicit</AssignmentMode>
  <AssignedInputType>Subject</AssignedInputType>
  <AssignedInputTag>ID</AssignedInputTag>
  <AssignedInputValue>01</AssignedInputValue>
  <AssignedOutputType>PhysiologicCollectionDevice</AssignedOutputType>
  <AssignedOutputTag>ID</AssignedOutputTag>
  <AssignedOutputValue>32</AssignedOutputValue>
</Assignment>
```

For each type of assignment, all the elements of the assignment object should be present and in the order presented here.

WPSM STUDY ARCHIVE

Directory Structure

Figure B5: WPSM Study Archive Directory Structure

Base Directory
Where
When
XML Tagged Data

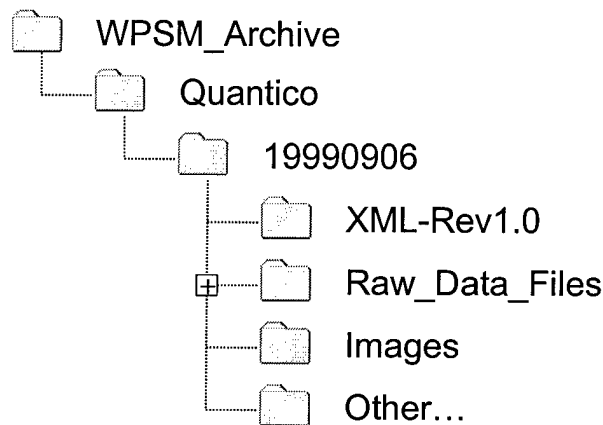


Figure B5 details the directory structure utilized by the data archive. Four directories are mandatory in the archive structure and are as follows: WPSM_Archive, Where, When, and XML. The WPSM_Archive directory separates the data archive from the root directory of the storage device. The Where directory level specifies test areas, training facilities, or any other location where a WPSM system is fielded. Within the

Where directory, the When directory level specifies the date when the experiment occurred. The XML directory contains the XML files for that study.

Any other directories under the When directory are optional. These may be used to store raw data from sensors or weather stations, or they may contain scanned copies of volunteer agreement affidavits. These directories may be referenced by data in the XML files, to show that further information is available and to allow these data items to be searched. Unless items or directories are referenced by data in the XML files, no automated searching is conducted on the optional files.

Directory Naming Convention

Where. Directory level names can be any combination of alphanumeric text. The only other characters permitted are underscores (_) and hyphens (-). The directory names are case sensitive and should represent the location where the WPSM system is being used.

When. Directory level names may only consist of numeric text in the form of a date. The date should be written with the first four characters representing the year; the next two, the month; and the last two, the day. For example, September 9, 1999, should be represented as the following:

19990909

If directories are sorted numerically, this will provide a structure in which the oldest date is first in the list of directories. When studies occur over extended periods of time, the starting date should be used for the directory name.

XML File Naming Convention

The standard file extension for an XML file is **.xml**.

Core File Name. Core file names will be made from a concatenation of the Where and When directory names, using an underscore between the two names. For example, if the Where directory name is "FtBenningDBBL" and the When directory name is "20000102," then the core file name would be the following:

FtBenningDBBL_20000102

This form immediately places the file in the correct location within the data archive. The When portion of the file name will change throughout the study.

XML Archive File Types. For organizational ease when defining study archives, files can be broken into three main categories: header, transient, and time series files.

Header Files contain all static data related to the study. These files would contain information and data that remain constant for the duration of the study.

Transient Files contain slow changing data, such as subject weight or pack weight. Slow changing data are defined as any parameter that changes every hour or less frequently. The header and transient files are named by appending “_Header” and “_Transient” to the core name, respectively. From the example above, the header and transient files’ full names name would be the following:

FtBenningDBBL_20000102_Header.xml
FtBenningDBBL_20000102_Transient.xml

Time Series Files contain data from parameters that change more frequently than every hour. These files contain the bulk of most WPSM study data. Depending on the size of the files and the amount of data, they can be broken down into periodic time chunks. The start hour of the time chunk should be appended to the base file name, such as the following:

FtBenningDBBL_2000010210.xml

All files in a WPSM archive can contain one or more than one XML object or data type. If a file contains data relating to only a few parameters, some indication of the data types represented should be appended to the base file name, such as the following:

FtBenningDBBL_2000010210_CoreTemp.xml

Other Directories in the Archive

Other directories in the archive are optional, but in most cases they will be necessary to fully document a study. There can be any number of other directories. Their purpose is to provide a structure to place raw data files, image files, scans of volunteer affidavits, and electronic copies of original documents. Files placed in these directories should be referenced within the XML files so raw data files can be traced back to individual subjects and queried. As with all XML records, the files in these directories contain data that relate to subjects and are valid for a portion or all of the study.

APPENDIX C
TAG DICTIONARY FOR U.S. MARINE CORPS INFANTRY OFFICER COURSE,
QUANTICO, VA, SEPTEMBER 1999

WPSM XML DATA TYPES

Table C1 shows the data type tags, their associated persistence, and the entity identifier.

Table C1: WPSM Data Types, Persistence, and Entities

Data Type Tag	Persistence	Entity ID	Description
<Desc>	@STUDY	@ALL	Descriptive Experimental Data
<Hh2o>	@STUDY	Subject	Doubly labeled water energy expenditure values
<S.h>	@STUDY	Subject	Subject Static Information
<S.t>	@STICKY	Subject	Subject Biographic Information
<Wld>	@STICKY	Subject	Subject Loaded Weight Information
<Clo>	@STICKY	Subject	Clothing Log
<Act>	@STICKY	Subject	Activity Log
<Eqp>	@STICKY	Subject	Equipment Log
<Food>	@STICKY	Subject	Food Log
<Met.loc>	@STICKY	Subject	Weather Station Location Information
<Met>	3600	W. Station	Meterological Information
<Phys>	Instant	Hub	Physiological Information
<Hub>	@STICKY	Hub	Hub Setup Information
<Image>	Variable	Variable	Study Picture
<FOOD_TABLE>	@STUDY	Food Type	Macro Nutrient Breakdown of Food Type
<CLOTHING_TABLE>	@STUDY	Clothing	Clo and Im Values for Clothing Types
<EQP_TABLE>	@STUDY	Equipment	Equipment weights for coded equipment
<ACTIVITY_TABLE>	@STUDY	Activity	Level of movement and activity level code

DATA TYPE ELEMENT DESCRIPTIONS

The data types and their complete set of elements are listed and described.

<Desc> represents descriptive information.

Persistence: @STUDY

Entity: @ALL

Data contained within this data type are valid for the whole study period, and thus they are placed in the header archive file. These data relate to the whole study and not to any one entity.

<u>Element Tag</u>	<u>Description</u>
<STFFVersion>	STFF protocol version number
<StudyTitle>	Study Title
<HURC>	HURC Number
<ProtocolNum>	Protocol Number
<NotebookNum>	Notebook Number
<SDT>	Start Date Time yyyyymmddhhnss.s format
<EDT>	End Date Time yyyyymmddhhnss.s format
<NorthLAT>	Northern Most Latitude of study
<SouthLAT>	Southern Most Latitude of Study
<WestLON>	Western Most Longitude of Study
<EastLON>	Eastern Most Longitude of Study
<LocationName>	Location or Locality Name
<USGSMap>	USGS Map number

<S.h> represents subject information.

Persistence: @STUDY

Entity: Subject

As these data are valid for the whole study, they are placed in the header file. Data contained within this data type are valid for the whole study period.

<u>Element Tag</u>	<u>Description</u>
<Last>	Last Name
<First>	First Name
<MI>	Middle Name
<SS>	Social Security Number
<DOB>	Date of Birth
<Age>	Age (Years)
<Add>	Street Address
<ANum>	Apartment Number
<City>	City
<ST>	State
<ZIP>	Zip Code
<PhHm>	Home Phone Number
<PhWk>	Work Phone Number
<Rank>	Rank (3 Letter)
<MOS>	Military Operational Specialty
<VAff>	Volunteer Affidavit Received (yes/no)
<VReg>	Volunteer Registry Data Sheet Received (yes/no)
<Sex>	Gender (M / F)
<Ht>	Height (Meters)
<Wt>	Weight Reported Weight (Kg)

<Hh2o> represents energy expenditure values obtained from doubly labeled water.

Persistence: @STUDY
Entity: Subject

These data are energy expenditure values averaged over 7 and 9 days. These data relate to the whole study period.

<u>Element Tag</u>	<u>Description</u>
<EE7DayAve>	Energy expenditure estimated over 7 days (Kcal/day)
<EE9DayAve>	Energy expenditure estimated over 9 days (Kcal/day)

<S.t> represents subject biographic information.

Persistence: @STICKY
Entity: Subject

Although these data are collected only twice, they belong in the transient file, as they do change within the study. These data were collected at the start and end of the study to examine change in body composition. A persistence of @STICKY has been chosen, as the values may change from time to time.

<u>Element Tag</u>	<u>Description</u>
<Wsn>	Weight Semi-nude (Kg)
<Wld>	Weight Load (Kg)
<Cnk>	Circumference (Neck) (cm)
<Cst>	Circumference (Stomach) (cm)
<BFat>	Body Fat (%)

<Wld> represents subject loaded weight.

Persistence: @STICKY
Entity: Subject

These data vary from day to day. Hence, they belong in the transient file. These data are collected every day at varying times. The data are valid until they are replaced by new data.

<u>Element Tag</u>	<u>Description</u>
<Wld>	Weight - Loaded (Kg)

<Hub> represents data collection hub configuration information.

Persistence: @STICKY
Entity: Hub Identifier

Although these data rarely change, hub configurations are nonetheless changeable and thus need to be in the transient data file. Note also when pills are changed, the HUB configuration also changes. A hub has a valid configuration until it is changed.

<u>Element Tag</u>	<u>Description</u>
<HubS>	Hub Serial Number

<code><sensor_type></code>	Sensor Serial Number / Sensor Number
<code><PillS></code>	HTI Pill Serial Number
<code><PillC></code>	HTI Pill Calibration Number

where *sensor_type* is the tag defined in data type `<Phys>` for physiologic sensors, such as the following:

<i>sensor_type</i>	
<code><1></code>	PCD Temperature Puck
<code><10></code>	BCTM

An example of a data object would be the following:

```

<Hub>
  <Persistence>@STICKY</Persistence>
  <Time>20000908123000</Time>
  <ID>24</ID>
  <S1>02</S1>
  <S10>30</S10>
</Hub>

```

This data object identifies the sensor types and the associated sensor identifiers, which were assigned to the hub from this point in time. The sticky persistence allows these data to be valid from this time on until information for this hub changes.

<Phys> represents physiologic information.

Persistence: Instant Data
 Entity: Hub Identifier

These data are collected every minute. This set of data is assumed to be valid for a point in time. It does not persist. Physiologic data are collected on a hub data collection device. Subjects can be assigned different hubs at any time during the study. Data are stored in the hub under a hub identifier. To link physiologic data to a subject, an explicit assignment needs to be generated whenever a subject receives a new hub.

<u>Element Tag</u>	<u>Description</u>
<1>	PCD Temperature Puck
<3>	PCD Heart Rate
<10>	Body Core Temperature Monitor (BCTM) Core Temp (°C)
<10.1>	BCTM Pill Present (1) / Absent (0)
<12>	Expended Energy Monitor (Calories)
<15>	PCD Wrist Temperature (Skin Side) (°C)
<16>	PCD Wrist Temperature (Outside) (°C)
<17>	PCD Sleep Score (minutes of sleep in last 24 hours)
<17.1>	PCD Sleep (minutes of sleep in last 15 minutes)
<18>	PCD Wrist Actigraphy (zero crossings per minute)
<18.1>	PCD Awake (0=false, 1=true)
<19>	PCD Chest Temp (Chest strap skin temp)
<20>	PCD Chest Strap Air Temp and Chest Actigraphy
<30>	PED Expended Energy Monitor (EEM) Even (Obsolete)
<31>	PED EEM Odd (Obsolete)
<32>	PED AMS (Ambulatory Monitoring System) Foot POD/EEM) Status
<32.1>	PED AMS Speed (miles per hour)
<32.2>	PED AMS Distance (total miles)
<32.3>	PED AMS Energy Expenditure due to Locomotion (total calories)
<33>	PED Heart Rate Logger (HRL) Status
<33.1>	PED HRL Average Heart Rate (bpm)
<33.2>	PED HRL Min Heart Rate (bpm)
<33.3>	PED HRL Max Heart Rate (bpm)
<33.4>	PED HRL Current Heart Rate (bpm)
<34.2>	PED GPS Unit Altitude
<34.3>	PED GPS Mode 1
<34.4>	PED GPS Mode 2
<34.5>	PED GPS Number of Satellites
<LAT>	PED GPS Latitude (Degrees)
<LON>	PED GPS Longitude (Degrees)

<Met> represents meteorologic information.

Persistence: 3540 (seconds)

Entity: Weather Station

These data are collected once per hour or more frequently, and are deemed valid for one hour. The persistence is set to 59 minutes to avoid data overlap.

MERCURY System Tags:

<u>Element Tag</u>	<u>Description</u>
<e>	Elevation
<T>	Air Temperature
<D>	Dew Point
<s>	Wind Speed
<d>	Wind Direction
<P>	Barometric Pressure
<x>	Solar Radiation
<g>	Globe Temperature
<G>	Ground Temperature
<r>	Relative Humidity

Smart Sensor Web, Weather Web Tags:

<u>Element Tag</u>	<u>Description</u>
<Tp>	Temperature (°C)
<Dp>	Dew Point Temperature (°C)
<Ws>	Average Wind Speed (ms ⁻¹)
<Wd>	Average Wind Direction (Degrees True North)
<Wg>	Wind Gust (ms ⁻¹)
<Pr>	Pressure (mb)
<Ca1>	Cloud Amount First Layer (eights)
<Ch1>	Cloud Amount First Layer (m)
<Ca2>	Cloud Amount Second Layer (eights)
<Ch2>	Cloud Amount Second Layer (m)
<Ca3>	Cloud Amount Third Layer (eights)
<Ch3>	Cloud Amount Third Layer (m)
<Vi>	Visibility (Km)
<Wx>	Present Weather Type (synoptic codes)

<Met.loc> represents weather station location information.

Persistence: @STICKY

Entity: Weather Station

Weather stations move only rarely. Weather station location data are valid until the weather station is moved.

<u>Element Tag</u>	<u>Description</u>
<LocationName>	Name of Location
<StationName>	Name of Weather Station
<LAT>	Latitude
<LON>	Longitude
<ALT>	Altitude / Elevation

<Act> represents activity log information.

Persistence: @STICKY

Entity: Subject

Warfighter activity data are collected every hour. Activity data are valid until updated.

<u>Element Tag</u>	<u>Description</u>
<Act>	Activity description.

Note 1: Activities are coded according to the scheme in Appendix D.

Note 2: For more than one activity, the <Act> tag is used multiple times with different datum.

<Clo> represents clothing log information.

Persistence: @STICKY

Entity: Subject

Clothing data are collected every hour. Clothing data are valid until updated.

<u>Element Tag</u>	<u>Description</u>
<Cloth>	Clothing Description

Note 1: Clothing items are coded according to the scheme in Appendix D.

Note 2: To describe many items of clothing, use the <Cloth> tag as many times as necessary.

<Eqp> represents equipment log information.

Persistence: @STICKY

Entity: Subject

Equipment data are collected every hour. Equipment data are valid until updated.

<u>Element Tag</u>	<u>Description</u>
<Eqp>	Equipment Item Description

Note 1: Equipment items are coded according to the scheme in Appendix D.

Note 2: To describe many items of equipment, use the <Eqp> tag as many times as necessary.

<Food> represents food information.

Persistence: @STICKY

Entity: Subject

Food data are collected every day. Food data are valid until updated.

<u>Element Tag</u>	<u>Description</u>
<Kcal>	Total Kcals Consumed
<Wt>	Total Weight of Food Consumed (Kg)
<Pro>	Total Weight of Protein Consumed (Kg)
<CHO>	Total Weight of Carbohydrate Consumed (Kg)
<Fat>	Total Weight of Fat Consumed
<Food>	Food Item Description

Note 1: Food items are coded to the scheme in Appendix D.

Note 2: To describe food items, use the <Food> tag as many times as necessary.

<Image> represents image information.

Persistence: Varies

Entity: Many

Images are collected at varying times. Image persistence is determined by the principle investigator. An image may or may not relate to an entity within the study. The entity may or may not be a subject. For example, in WPSM, a picture of a weather station relates to weather stations but not subjects. Where an image relates to many subjects or entities, use separate XML image objects to relate the image to those entities.

<u>Element Tag</u>	<u>Description</u>
<Name>	Image Name
<KeyWord> ¹	Key Words to Describe Picture
<PFormat>	File Type (e.g., JPEG)
<Pid>	Relative Path to Image File in Archive, with File Name and Extension (e.g., Images\MarineCleaningWeapon.bmp).

¹ For more than one key word, use this tag more than once.

<FOOD TABLE> references all the possible food types encoded during this study and provides additional information on each.

Persistence: @STUDY

Entity: Food Type Identifier (Listed in <Food> data object)

These data are valid for the entire study.

<u>Element Tag</u>	<u>Description</u>
<Kcal>	Energy of Portion (KCal)
<Wt>	Weight (g)
<Pro>	Total Weight of Protein in Portion (g)
<CHO>	Total Weight of Carbohydrate in Portion (g)
<Fat>	Total Weight of Fat in Portion (g)

<CLOTHING TABLE> references all the possible clothing types encoded during this study and provides additional information on each.

Persistence: @STUDY

Entity: Clothing Type Identifier (Listed in <Clo> data object)

These data are valid for the entire study.

<u>Element Tag</u>	<u>Description</u>
<Wt>	Clothing Weight (Kg)
<Clo>	clo Value (clo)
<Im>	Clothing Permeability Index (Im)

<EQP TABLE> references all the possible equipment types encoded during this study and provides additional information on each.

Persistence: @STUDY

Entity: Equipment Type Identifier (Listed in <Eqp> data object)

These data are valid for the entire study.

<u>Element Tag</u>	<u>Description</u>
<Wt>	Equipment Weight

<ACTIVITY TABLE> references all the possible activities encoded during this study and provides additional information on each.

Persistence: @STUDY

Entity: Activity Type Identifier (Listed in <Act> data object)

These data are valid for the entire study.

<u>Element Tag</u>	<u>Description</u>
<Level_of_Movt>	Level of Movement for the Given Activity
<Act_Level>	Coded Activity Level

APPENDIX D
CODING SCHEMES FOR ACTIVITY, CLOTHING, EQUIPMENT, AND FOOD FOR
U.S. MARINE CORPS INFANTRY OFFICER COURSE,
QUANTICO, VA, SEPTEMBER 1999

ACTIVITY CODING SCHEME

Among the data collected from participants of the USMC Infantry Officer Course Cold Weather Field Exercise, conducted in Quantico, VA September 1999, was an Activity Log. Due to the open-ended textual nature of the Activity Log entries, responses were summarized and evaluated for commonality. The Activity Log responses were then collapsed into appropriate activity-oriented categories, based on the nature of the log entry, as well as the objectives/itinerary of the Cold Weather field Exercise. All appropriate activity-oriented categories were then coded (see below), using 4 alpha letters to denote each activity. The following is a list of coded activities:

- Administration (ADMN)
- Ambush (AMBH)
- Ammunition Turn-In (AMMO)
- Anti-Mechanized (ANME)
- Assembly Area (ASAR)
- Attack (ATCK)
- Barrier Plan (BRPL)
- Battle Position (BATL)
- Break (BREK)
- Brief (BREF)
- Camouflage (CAMO)
- Classes (CLAS)
- Combat (CBAT)
- Convoy (CVOY)
- Debrief (DBRF)
- Defense (DEFN)
- Digging-In (DGIN)
- Eat (EATT)
- Firewatch (FIRW)
- Firing (FIRE)
- Foot Movement (FOOT)
- Gear Adjustment (GEAR)
- Grenade (GRND)
- Helicopter Extraction (HELC)
- Helicopter Movement (HELO)
- Infiltration (INFL)
- Insert (INST)
- Inspection (INSP)
- Live Fire (LFIR)
- Load Zone (LDZN)

Maintenance (MAIN)
Mechanized Movement (MECH)
Mechanized/Motorized Movement (MEMO)
Mechanized/Motorized/HELO Movement (MEMO)
Meeting (MEET)
Mortar (MORT)
Movement (MOVT)
NO LOG (NLOG)
Objective Rally Point (ORPP)
Organized (ORGN)
Orders (ORDR)
Pack-Up (PACK)
Patrol (PTRL)
Person Hygiene (PHYG)
Planning (PLAN)
Position (PSTN)
Radio-Watch (RDIO)
Range (RNGE)
Reconnaissance Patrol (RCON)
Rehearsal (REHR)
Relief In Place (RELF)
Research Team (RSCH)
Rest (REST)
Re-supply (RSUP)
Reveille (REVL)
Rucksack (RUCK)
Sat (SATT)
Sleep (SLEP)
SP/LP (SPLP)
Staging (STAG)
Stand To (STAN)
Stationary (STNY)
Strongpoint (SRNG)
Terrain Model (TMOD)
Training (TRNG)
Truck (TRUK)
UNSPECIFIED (UNSP)
Wait (WAIT)
Warning Order (WARN)
Watch (WTCH)
Weapons (WPNS)
Weight-In (WGHT)

Application of the Activity Log Coding Scheme

The application of the activity log coding scheme includes replacing the original Activity Log entries with a response from the Activity Log Coding Scheme. Each cell will contain one (or more) Activities (coded), separated by a backslash and then a numeric code (0-5) for Type of Movement (see below) side-by-side with a numeric code (0-9) for Activity Level (see below). Some of the Activity (coded) responses have additional relevant information enclosed in brackets (see below).

Type of Movement:

- 0 = UNSPECIFIED
- 1 = Foot Movement
- 2 = Mechanized Movement
- 3 = Mechanized/Motorized Movement
- 4 = Mechanized/Motorized/HELO Movement
- 5 = Stationary

Activity Level Scale: (0-7)

Scale anchors:

- 0 = None
- 2 = Low
- 4 = Mod
- 6 = High

Examples of Coded Activity:

ATCK/16 = Attack with foot movement with a high level of activity.

MOVT (VIA 5-TON)/22 = Movement via 5 ton truck, mechanized with a low level of activity.

Coding Applied to Activity Log Entries

Table D1 applies the coding scheme to all reported activities.

Table D1: Reported and Coded Activities

ACTIVITY LOG RESPONSE	ACTIVITY (CODED)	LEVEL OF MOVEMENT 0 = UNSPECIFIED (UNSP) 1 = Foot Movement (FOOT) 2 = Mechanized Movement (MECH) 3 = Mechanized/Motorized Movement (MEMO) 4 = Mechanized /Motorized/ HELO Movement (HEMO) 5 = Stationary (STNY)	ACTIVITY LEVEL 0 = UNSPECIFIED (UNSP) 2 = Low 4 = Mod 6 = High
Ambush	AMBH	1	6
Ambush (Execute)	AMBH(EXECUTE)	1	5
Ammo (Turn-In)	AMMO(TURN-IN)	1	2
Ammo (Draw)	AMMO(DRAW)	5	2
Assembly Area (Occupy)	ASAR(OCCUPY)	1	2
Attack	ATCK	1	6
Attack (240)	ATCK(240)	1	6
Attack (Counter)	ATCK(COUNTER)	1	6
Attack (Mechanized)	ATCK(MECHANIZED)	2	6
Attack (NBC)	ATCK(NBC)	1	4
Attack (Night)	ATCK(NIGHT)	1	6
Attack (On LZ Bluejay - Mechanized)	ATCK(ON_LZ_BLUEJAY_MECHANIZED)	2	6
Attack (Prepare For)	ATCK(PREPARE_FOR)	2	4
Barrier Plan	BRPL	1	3
Barrier Plan (Organized)	BRPL(ORGANIZED)	1	2
Battle Position (Occupy)	BATL(OCCUPY)	1	2
Break	BREK	5	1
Break (At R-15)	BREK(AT_R-15)	5	1
Brief	BREF	5	1
Cammo Paint	CAMO	5	1
Classes	CLAS	5	1
Combat (Prepare For)	CBAT(PREPARE_FOR)	1	3
Convoy (Prepare For)	CVOY(PREPARE_FOR)	1	3
Convoy (Tactical)	CVOY(TACTICAL)	2	2
Debrief (Of)	DBRF	5	1
Defense (Occupy)	DEFN(OCCUPY)	1	2
Defense (Of)	DEFN	1	5
Defense Of Attack	DEFN(OF_ATTACK)	1	5
Defensive Position/Defense Posture	DEFN	1	2
Eat	EATT	5	2
Faked Injury	TRNG	5	2

Fighting Holes - Digging In	DGIN	1	5
Firewatch	FIRW	1	2
Firewatch (Registration)	FIRW(REGISTRATION)	1	1
Firing Position/Firing	FIRE	1	2
Gear Adjustments	GEAR	1	3
Grenade	GRND	1	0
Grenade (Range)	GRND(RANGE)	1	5
Grenade (Throw)	GRND(THROW)	1	5
Helicopter Extract	HELCO	1	5
Infiltration	INFL	1	6
Insert (Into LZ)	INST(INTO_LZ)	4	5
Inspection	INSP	5	2
Live Fire	LFIR	1	6
Live Fire (Anti-Mechanized)	LFIR(ANTI-MECHANIZED.)	5	4
Live Fire (At-4)	LFIR(AT-4)	1	4
Live Fire (Prepare For)	LFIR(PREPARE_FOR)	1	3
LZ (Arrive At)	LDZN(ARRIVE_AT)	2	2
LZ (Bluejay)	LDZN(BLUEJAY)	2	2
LZ (Picked-Up From Via Helicopter)	LDZN(PICKED-UP_FROM_VIA_HELOCOP TER)	4	4
Meeting (With Research Team)	MEET(RSCH)	5	1
Mortar Drills	MORT	1	4
Movement	MOVT	1	4
Movement (For Infiltration)	MOVT(FOR_INFILTRATION)	1	4
Movement (From Attack)	MOVT(FROM_ATTACK)	1	4
Movement (From Exfiltration)	MOVT(FROM_EXFILTRATI ON)	1	4
Movement (From R-11)	MOVT(FROM_R-11)	2	4
Movement (From R-15)	MOVT(FROM_R-15)	1	4
Movement (Night)	MOVT(NIGHT)	1	4
Movement (To Ambush Site)	MOVT(TO_AMBUSH_SITE)	1	4
Movement (To Anti-Mechanized Range)	MOVT(TO_ANTI-MECH_RANGE)	1	4
Movement (To Assault Green)	MOVT(TO_ASSAULT_GRE EN)	1	4
Movement (To Assembly Area)	MOVT(TO_ASSEMBLY_AR EA)	1	4
Movement (To Attack)	MOVT(TO_ATTACK)	1	4
Movement (To Combat Town)	MOVT(TO_COMBAT-TOWN)	2	2

Movement (Walk To Combat Town)	MOVT(WALK_TO_COMBAT-TOWN)	1	4
Movement (To Consolidate)	MOVT(TO_CONSOLIDATE)	1	4
Movement (To Defense Position)	MOVT(TO_DEFENSE_POSITION)	2	4
Movement (To Get Rucks)	MOVT(TO_GET_RUCKS)	1	4
Movement (To Grenade Range)	MOVT(TO_GRENADE_RANGE)	1	4
Movement (To Listening Post)	MOVT(TO_LISTENING_POST)	1	4
Movement (To LZ Bluejay)	MOVT(TO_LZ_BLUEJAY)	1	4
Movement (To LZ Chickadee)	MOVT(TO_LZ_CHICKADEE)	1	4
Movement (To LZ Finch)	MOVT(TO_LZ_FINCH)	1	4
Movement (To LZ)	MOVT(TO_LZ)	1	4
Movement (To LZ-19)	MOVT(TO_LZ-19)	1	4
Movement (To LZ-21)	MOVT(TO_LZ-21)	1	4
Movement (To Next Defensive Position)	MOVT(TO_NEXT_DEFENSIVE_POSITION)	1	4
Movement (To Objective)	MOVT(TO_OBJECTIVE)	1	4
Movement (To ORP)	MOVT(TO_ORP)	1	4
Movement (To R-11)	MOVT(TO_R-11)	1	4
Movement (To R-15 Treeline)	MOVT(TO_R-15_TREELINE)	1	4
Movement (To R-15 Via 5-Ton)	MOVT(TO_R-15_VIA_5-TON)	2	2
Movement (To R-3)	MOVT(TO_R-3)	2	4
Movement (To R-3B On Bus)	MOVT(TO_R-3B_VIA_BUS)	2	2
Movement (To R-8)	MOVT(TO_R-8)	1	4
Movement (To Range 3-A)	MOVT(TO_RANGE_3-A)	1	4
Movement (To Range Via Bus)	MOVT(TO_RANGE_VIA_BUS)	2	2
Movement (To Range)	MOVT(TO_RANGE)	1	4
Movement (To Sleep Site)	MOVT(TO_SLEEP_SITE)	1	4
Movement (To Strongpoint)	MOVT(TO_STRONGPOINT)	1	4
Movement (Via 5-Ton)	MOVT(VIA_5-TON)	2	2
Movement (Via Truck)	MOVT(VIA_TRUCK)	2	2
Movement (To Woodline)	MOVT(TO_WOODLINE)	1	4
No Log	NLOG	N/A	N/A
Orders (Issued)	ODRS(ISSUED)	5	2
Orders (Prepared)	ODRS(PREPARED)	5	2
Orders (Processed)	ODRS(PROCESSED)	5	2

Orders (Received)	ODRS(RECEIVED)	5	2
Organize	ORGN	1	2
ORP (Occupied)	ORPP(OCCUPIED)	1	3
Pack-Up	PACK	1	4
Patrol	PTRL	1	4
Personal Hygiene	PHYG	5	2
Planning	PLAN	5	2
Position (Establish)	PSTN(ESTABLISH)	1	3
Radio-Watch	RDIO	1	2
Range Walk	RNGE	1	3
Reconnaissance Patrol	RCON	1	4
Reconnaissance Patrol (Battle Position)	RCON(BATTLE_POSITION)	1	4
Reconnaissance Patrol (Combat)	RCON(COMBAT)	1	4
Reconnaissance Patrol (D)	RCON(D)	1	4
Reconnaissance Patrol (MG Position)	RCON(MG_POSITION)	1	4
Reconnaissance Patrol (Range)	RCON(RANGE)	1	4
Reconnaissance Patrol (New Position)	RCON(NEW_POSITION)	1	4
Reconnaissance Patrol (Objective)	RCON(OBJECTIVE)	1	4
Rehearsal	REHR	1	4
Rehearsal (For Mechanized Attack)	REHR(FOR_MECHANIZED_ATTACK)	2	4
Rehearsal (Radio-Watch)	REHR(FOR_RADIO-WATCH)	1	3
Relief In Place	RELF	5	1
Research Team (Gear Set-Up)	RES(SET-UP)	1	2
Research Team (Mtg.)	RES(MEETING)	5	2
Rest	REST	5	1
Resupply	RSUP	5	3
Reveille	REVL	5	2
Rucksack	RUCK	0	2
Rucksack (Load)	RUCK(LOAD)	5	2
SP/Listening Post	SPLP	1	2
SP/LP (Occupy)	SPLP(OCCUPY)	1	3
SP/LP (Manning)	SPLP(MANNING)	1	3
Sat In Site	SATT	5	2
Sleep	SLEP	5	1
Staging	STAG	1	3

Stand To	STAN	1	2
Strongpoint	SRNG	1	4
Strongpoint (Occupy)	SRNG(OCCUPY)	1	3
Terrain Model Created/Modified	TMOD	1	2
Terrain Walk	TMOD	1	4
Training Classes	TRNG	5	1
Truck (Load)	TRUK(LOAD)	1	4
Waiting	WAIT	1	2
Waiting (For Attack)	WAIT(FOR_ATTACK)	1	2
Warning Order (Prepare)	WARN(PREPARE)	1	1
Watch	WTCH	5	2
Weapon (Firing)	WPNS(FIRING)	1	4
Weapons (Maintenance)	WPNS(MAINTENANCE)	5	2
Weigh-In	WGHT	5	2

CLOTHING CODING SCHEME

Among the data collected from participants of the USMC Infantry Officer Course Cold Weather Field Exercise, Quantico, VA, September 1999, was a Clothing Log. Due to the open-ended textual nature of the Clothing Log entries, responses were summarized and evaluated for commonality. The Clothing Log responses were then collapsed into appropriate combat clothing-oriented categories, based on the following document: Combat Clothing and Equipment (U.S. ARMY SOLDIER AND BIOLOGICAL CHEMICAL COMMAND, MARCH 1998, Natick P-32-1). All appropriate clothing-oriented categories were then coded (see Table D2), using alpha letters to denote each clothing category.

The following is a list of clothing log entries:

- Boot--Combat, Mildew and Water Resistant, Direct Molded Sole
- Coat--Uniform, Battledress, Temperate Zone
- Drawers--Underwear, Extended Cold Weather, Polypropylene
- Helmet--Ground Troops and Parachutist's (with Parachute Pads)
- Jacket--Field
- Parka--Wet Weather
- Poncho--Wet Weather, Camouflage
- Socks--Men's, Wool, Cushion Sole, Stretch Type
- Trousers--Uniform, Battledress, Temperate Zone
- Trousers--Wet Weather
- Undershirt--Cotton
- Underwear--Undershirt, Extended Cold Weather, Polypropylene
- Vest--Individual, Tactical Load Bearing
- Vest--Body Armor, Fragmentation Protective Vest, Personnel Armor System for Ground Troops (PASGT)
- Load Bearing Equipment--Belt and Suspenders, Individual Equipment
- MOPP Gear--Coat and Trousers
- MOPP Gear--Gloves
- MOPP Overboots

Application of the Clothing Log Coding Scheme

The application of the clothing log coding scheme includes replacing the original Clothing Log entries with a response from the Clothing Log Coding Scheme. Each cell will contain one coded clothing item, separated by a backslash and then a number (ranging from 0.00-1.00) to denote Clothing Insulation (CLO) value (see below), separated by a backslash and then a number (xx.xx)--the clothing item's weight (in Kilograms) (see below). Although some researchers values include a Clothing Permeability (Im) and index of evaporative loss, with CLO value, such values will not be incorporated into this project due to a lack of information.

Example of Coded Clothing:

PASGTH/1.42/1.50 = Ground troop helmet with CLO of 1.42 and weight of 1.5 Kg

Several of the clothing codes were grouped into clothing ensembles (see below). The purpose of using clothing code ensembles is to reduce the necessary data for the Clothing variable final data set. The Ensemble (Clothing code, BDUENS) contains the following combat clothing items:

- Boot--Combat, Mildew and Water Resistant, Direct Molded Sole
- Socks--Men's, Wool, Cushion Sole, Stretch Type
- Trousers--Uniform, Battledress, Temperate Zone
- Coat--Uniform, Battledress, Temperate Zone

Coding for the Quantico, VA, September 1999 clothing log entries is shown in Table D2.

Table D2: Reported and Coded Clothing

CLOTHING LOG RESPONSE	CLOTHING (CODED)	CLO VALUE	WEIGHT (KG.)
Boot--Combat, Mildew and Water Resistant, Direct Molded Sole	CBTBT	1.20 (includes Socks, Men's, Wool, Cushion Sole, Stretch Type)	1.86 kg per pair, size 9
Coat--Uniform, Battledress, Temperate Zone	TZCOAT		1.43 kg - includes Coat, Uniform, Battledress, Temperate Zone
Drawers--Underwear, Extended Cold Weather, Polypropylene	POLYPRO		0.37 kg
Helmet--Ground Troops and Parachutist's (with Parachute Pads)	PASGTH	1.42	1.50 kg
Jacket--Field	JKT		
Parka--Wet Weather	PARKA		1.05 kg - includes Trousers, Wet Weather
Poncho--Wet Weather, Camouflage	PONCHO		0.68 kg
Socks--Men's, Wool, Cushion Sole, Stretch Type	SOX	0.75	0.8 kg /pr

Trousers--Uniform, Battledress, Temperate Zone	TZTRS		1.43 kg - includes Trousers, Uniform, Battledress, Temperate Zone
Trousers--Wet Weather	WWTRS		1.05 kg - includes Parka, Wet Weather
Undershirt--Cotton	USHRT		0.08 kg
Underwear--Undershirt, Extended Cold Weather, Polypropylene	POLYSHT		0.31 kg
Vest--Individual, Tactical Load Bearing	TACLBV		0.9 kg
Vest--Body Armor, Fragmentation Protective Vest, Personnel Armor System for Ground Troops (PASGT)	PASGTV		4.09 kg - Size Medium
Load Bearing Equipment--Belt and Suspenders, Individual Equipment	LBE		
MOPP Gear--Coat and Trousers	MOPP		
MOPP Gear--Gloves			
MOPP Overboots			
ENSEMBLES:			
Boot--Combat, Mildew and Water Resistant, Direct Molded Sole; AND	BDUENS	1.53	4.08 kg
Socks--Men's, Wool, Cushion Sole, Stretch Type; AND			
Trousers--Uniform, Battledress, Temperate Zone; AND			
Coat--Uniform, Battledress, Temperate Zone			
Trousers--Wet Weather; AND	WWGEAR	1.64	3.58 kg
Parka--Wet Weather.			

EQUIPMENT CODING SCHEME

Among the data collected from participants of the USMC Infantry Officer Course Cold Weather Field Exercise, Quantico, VA, September 1999, was an Equipment Log. Due to the open-ended textual nature of the Equipment Log entries, responses were summarized and evaluated for commonality. The Equipment Log responses were then collapsed into appropriate military combat equipment oriented categories, based on, in part, the following document: Combat Clothing and Equipment (U.S. Army Soldier and Biological Chemical Command, March 1998, Natick P-32-1). All appropriate equipment-oriented categories were then coded (see Table D3), using alpha letters alone or alpha letters plus numeric digits to denote each equipment-oriented category.

Application of the Equipment Log Coding Scheme

The application of the equipment log coding scheme includes replacing the original Equipment Log entries with a response from the Equipment Log Coding Scheme. Each cell will contain one coded equipment item, separated by a backslash and then a number (xx.xx) representing equipment item weight in kilograms.

Example of Coded Clothing:

SAW/7.05 = Squad Automated Weapon, with a weight of 7.05 Kg

Coding for the Quantico, VA, September 1999, equipment log entries is shown in Table D3.

Table D3: Reported and Coded Equipment

EQUIPMENT LOG RESPONSE	EQUIPMENT (CODED)	WEIGHT (KG.)
M16	M16	3.60 kg
M60	M60	28.87 kg
Squad Automatic Weapon	SAW	7.05 kg
M-203 Attachment (for M-16)	M203	1.05 kg
PAC-4	PAC4	0.45 kg
Spare barrel for M-249	BARL	
A-bag	ABAG	25 kg
Night Vision Goggles	NVG	0.31 kg
Radio	RADIO	

FOOD TABLE

Table D4 shows the food categories used for the Quantico, VA, September 1999.

Table D4: Food Table

Food Item	Weight (Kg)	Kcal	Protein (Kg)	CHO (Kg)	Fat (Kg)
Cream Substitute, Powder	2.0	10.93	0.10	1.10	0.71
Ergo Drink Lemon Flavor	47.0	170.00	0.00	43.00	0.00
ERGO Drink Tropical Punch Flavor	47.0	170.00	0.00	43.00	0.00
Fruit Leather	30.0	90.00	1.00	21.00	0.00
KELLOGG'S NUTRI-GRAIN Apple Cinnamon Cereal Bar	37.0	140.38	2.01	27.07	3.01
KELLOGG'S NUTRI-GRAIN Raspberry Cereal Bar	37.0	140.00	2.00	27.00	3.00
KELLOGG'S NUTRI-GRAIN Strawberry Cereal Bar	37.0	140.00	2.00	27.00	3.00
MRE APPLESAUCE	128.0	88.00	0.36	22.83	0.09
MRE BEEF MUSHROOMS	227.0	312.00	27.92	11.39	16.90
MRE Beef Ravioli	227.0	300.00	18.00	37.00	8.00
MRE BEEF SNACKS	23.0	65.00	7.87	0.38	3.44
MRE BEEF STEW	227.0	242.00	27.71	12.55	8.45
MRE BEEF TERIYAKI	227.0	313.00	20.57	23.49	15.10
MRE BEV BSE DRY	34.0	132.00	0.04	33.85	0.02
MRE BREAD WHITE	51.0	51.00	4.58	26.41	6.04
MRE BROWNIE FUDGE	85.0	340.00	4.69	46.57	17.45
MRE CHARMS	28.0	111.00	0.01	27.60	0.04
MRE CHEESE SPREAD	4.0	171.00	5.53	0.67	16.51
MRE Cheese Tortellini in Tomato Sauce	227.0	170.00	5.00	31.00	2.50
MRE CHICKEN BREAST	79.0	115.00	15.88	2.25	4.19
MRE CHICKEN CAVETELLI	227.0	313.00	17.35	30.02	13.56
MRE CHICKEN NOODLE	227.0	207.00	17.12	17.01	7.39
MRE CHICKEN RICE	227.0	292.00	33.79	12.88	10.66
MRE CHICKEN SALSA	227.0	159.00	20.11	11.10	3.55
MRE CHICKEN STEW	227.0	260.00	21.18	18.13	11.16
MRE CHILI MAC	227.0	273.00	22.67	22.64	10.10
MRE CHOCOLATE/MINT POUND CAKE	71.0	293.00	4.15	40.29	14.53
MRE CHOW MEIN NOODLES	28.0	133.00	3.75	16.79	5.89
MRE CIDER MIX	19.0	73.00	0.12	18.68	0.05
MRE COCOA DRY	227.0	283.00	17.60	16.85	16.40
MRE COOKIE CHCV	43.0	220.00	3.02	26.68	11.96
MRE CRACKERS	19.0	88.00	1.87	13.23	2.74
MRE CREAM SUBS	4.0	22.00	0.19	2.15	1.42
MRE FRANKS	108.0	274.00	15.90	2.83	21.64
MRE FRUIT MIX WET	128.0	94.00	0.57	24.29	0.06
MRE GRANOLA BAR	45.0	208.00	4.11	30.69	8.20
MRE HAM SLICE	113.0	145.00	23.42	0.08	4.93
MRE JAM	28.0	69.00	0.20	18.26	0.06
MRE JELLY	28.0	77.00	0.11	20.10	0.03
MRE LEMON POUND CAKE	71.0	304.00	3.76	40.55	14.24

MRE M&M PLAIN	43.0	202.00	2.67	29.21	9.29
MRE MEATLOAF GRAVY	227.0	283.00	17.60	16.85	16.40
MRE MEXICAN RICE	142.0	194.00	4.35	35.56	3.94
MRE NOODLES IN BUTTER	142.0	187.00	3.35	17.96	0.00
MRE ORANGE POUND CAKE	71.0	307.00	3.60	40.63	14.69
MRE PEACHES WET	128.0	102.00	0.67	26.20	0.09
MRE PEANUT BUTTER	21.5	174.00	5.84	3.73	10.52
MRE PEANUTS	28.0	161.00	7.38	6.33	13.09
MRE PEARS WET	128.0	98.00	0.34	25.53	0.08
MRE PINEAPPLE POUND CAKE	71.0	306.00	3.61	40.92	14.34
MRE PINEAPPLE WET	128.0	92.00	0.47	23.96	0.08
MRE PORK CHOP	227.0	283.00	22.14	18.23	13.15
MRE PORK CHOW MEIN	227.0	224.00	16.37	13.63	11.53
MRE POTATO STICK	28.0	147.00	1.94	13.74	10.34
MRE PRETZELS	28.0	108.00	2.58	22.45	0.99
MRE SALT	4.0	0.00	0.00	0.00	0.00
MRE SHORTBREAD	43.0	216.00	2.78	27.75	10.31
MRE SKITTLES	62.0	241.00	0.40	56.34	2.87
MRE SPAG MT SCE	227.0	321.00	17.35	22.28	18.11
MRE Spiced Apples	142.0	150.00	0.00	34.00	2.50
MRE Strawberry Jam	28.0	80.00	9.00	20.00	0.00
MRE SUGAR	6.0	14.00	0.00	6.00	0.00
MRE TABASCO SAUCE	4.0	0.00	0.03	0.04	0.02
MRE TEA DRY	1.0	4.00	0.11	0.00	0.00
MRE THAI CHICKEN	227.0	217.00	15.87	13.85	10.86
MRE Toaster Pastries	52.0	200.00	2.00	37.00	5.00
MRE TOOTSIE ROLL	28.0	115.00	0.43	23.23	2.86
MRE TURKEY GRAVY	227.0	202.00	23.75	15.16	4.52
MRE VANILLA POUND CAKE	71.0	308.00	3.49	40.34	14.90
MRE WHEAT BREAD	51.0	166.00	4.11	26.25	5.17
MRE WHITE RICE	142.0	243.00	4.36	34.19	9.61
SNICKERS Bar	40.3	230.00	6.00	17.00	15.00

APPENDIX E: WPSM QUERY SERVER TEXT INTERFACE STANDARD

Communication with the WPSM Query Server uses four files: GQLCommand.txt, GQLQueryString.txt, GQLStudySummary.txt, and GQLQueryResults.txt.

GQLCommand.txt

All interactions with the server should start by writing an appropriate command file. The query server looks for the presence of this command file within its root directory. If present, the server executes the commands and then deletes the file.

The file "GQLCommand.txt" is an ASCII file containing two lines:

```
command
directory_string
```

Valid commands currently are "Rebuild," which causes GQLServer to read the specified archive and regenerate the GQLArchiveSummary.txt file, and "QueryType1," which tells GQLServer to look at the file GQLQueryString.txt.

The directory_string contains the full path to the data, including drive letters and backslashes (e.g., "c:\~wpsm\quantico_19999\19990906\cleaned_xml");

GQLQueryString.txt

The ASCII file "GQLQueryString.txt" contains the following data:

```
Granularity in seconds
StartGranuleIndex
EndGranuleIndex
PersistenceFlag--choices are "UsePersistence" or "DontUsePersistence"
GranularOperationFlag--choices are "GranuleFirst," "GranuleLast,"
    "GranuleMin," "GranuleMax," "GranuleAverage," or "GranuleAll"
QueryString--the query string output by DVDM client (e.g., "Phys.HeartRate >
    30")
Select TypeTag--a command to tell GQLServer which of the query results to
    actually return (e.g., "Select Phys.HeartRate". One "Select" string per line)
```

A sample GQLQueryString.txt file is below:

```
Granularity 60
0
15696
UsePersistence
GranuleAverage
{S.h}.{id}=01&{S.h}.{DOB}?&{S.h}.{Age}?&{S.h}.{Ht}?&{S.h}.{Wt}?&{S.t}.{BFat}?
```

Select {S.h}.{Age}
Select {S.h}.{DOB}
Select {S.h}.{Ht}
Select {S.h}.{Wt}
Select {S.t}.{BFat}

GQLStudySummary.txt

The ASCII file "GQLStudySummary.txt" is created after the archive is rebuilt using the "Rebuild" command in the GQLCommand file. It contains information on the archive, including the number of static, transient, and temporal lines, longest line length, number and names of data types in an archive, and for each data type, the enumerated IDs for this tag and the number and names of all of the data tags. All assignments are also listed in the file. GQLStudySummary.txt is read by the server when the archive is selected in the GUI by the client. A sample GQLStudySummary.txt file is below:

1 static lines
1 transient lines
198026 temporal lines
0 corrupt lines
83 longest line
20010630120000.00 timeMinSTF
20010706000000.00 timeMaxSTF
993916800 start secondsSince1970
994392000 end secondsSince1970
1 types
type 0 {Phys} 4 tags
25 IDs for this tag
0 {id} 03
1 {id} 04
2 {id} 05
3 {id} 06
4 {id} 07
5 {id} 08
6 {id} 09
7 {id} 10
8 {id} 11
9 {id} 12
10 {id} 13
11 {id} 14
12 {id} 15
13 {id} 17
14 {id} 18
15 {id} 19
16 {id} 20
17 {id} 21

18 {id} 22
19 {id} 23
20 {id} 24
21 {id} 25
22 {id} 26
23 {id} 27
24 {id} 01
0 {id}
1 {ProportionalData}
2 {ZCMDData}
3 {TATData}
0 associations

GQLQueryResults.txt

The ASCII file "GQLQueryResult.txt" contains the results of the query passed from the DVDM client to GQLServer. After the self-explanatory header, fields on the line include the following:

granuleIndex
Seconds Since Jan 1 1970
STFTime
ID Type (after assignment)
ID Tag (after assignment)
ID Value (after assignment)
Query Type
Query Tag
Query Data Type (i.e., INTEGER, FLOAT, etc.)
Query Value
Index ID Type
Index ID
Index Query Type
Index Query Tag
Index Data Type

GQLQueryResult.txt file is for debugging purposes only; it is not actually read by the DVDM client.

A sample GQLQueryResult.txt file follows:

2,nLines
15,nColumns
FirstGranule,0,LastGranule,1
0,936676800,19990907000000,{S.h},{id},01,{S.h},{id},INTEGER,01,1,0,1,0,2
0,936676800,19990907000000,{S.h},{id},02,{S.h},{id},INTEGER,02,1,1,1,0,2

GQLQueryResults.bin

The file "GQLQueryResult.bin" contains the same information as GQLQueryResult.txt but in a binary, non-ASCII format to conserve space and increase the speed with which the file is written and read. It is read by the DVDM Client.

PROCESS OF QUERYING A WPSM XML ARCHIVE USING THE QUERY SERVER

1) Direct Query Server to XML Archive

Write the file GQLCommand.txt in the root directory of the Query Server. The first line should be the command "Rebuild," and the second line should be the directory_string which is the full path to the XML archive.

2) Read GQLStudySummary.txt File

This file identifies the data type and entity ID which exist in the archive along with the time bounds of when the study took place. Queries that are sent to the Query Server should be based upon the output of this file.

3) Direct Query Server to Look at File GQLQueryString.txt

Write the file GQLCommand.txt with the command QueryType1 and the directory_string of the full path of the XML archive.

4) Write Query to GQLQueryString.txt

Write and save a file of the format described in the above section documenting GQLQueryString.txt.

5) Read Query Results in GQLQueryResults.txt

BIBLIOGRAPHY

The following bibliographic references are categorized into five main areas:

Temporal Databases (Temporal)
Bioinformatics/Computational Biology (Bio-Informatics)
Medical Informatics (Med-Infomatics)
Information Broker Architecture (Broker)
Software Design/Programming Languages (Software)

Abiteboul, S., L. Herr, and J. Bussche. Temporal connectives versus explicit timestamps to query temporal databases. *Journal of Computer and System Sciences* 58(1): 54, 1999. (Temporal)

Adiba, M., and M. J. L. Zechinelli. Spatio-temporal multimedia presentations as database objects. *Lecture Notes in Computer Science*, 1677: 1999. (Temporal)

Advani, A., et al. Integrating a modern knowledge-based system architecture with a Legacy VA Database: The ATHENA and EON Projects at Stanford. *AMIA [SMI*-1999-0795]*, 1999. (Broker)

Alarcon, P .A., A. Gupta, and J. M. Carazo. A framework for querying a database for structural information on 3D images of macromolecules: A web-based query-by-content prototype on the BiImage macromolecular server. *J. Struct. Biol.* 125(2-3): 112-22, 1999. (Bio-Informatics)

Altinel, M., D. Aksoy, T. Baby, M. Franklin, W. Shapiro, and S. Zdonik. DBIS-Toolkit: Adaptable Middleware for Large Scale Data Delivery, Demo Desc. for ACM SIGMOD Conference, PA, 1999. (<http://www.cs.umd.edu/~altinel/sigmod99/sigmod99>). (Broker)

AMIA Annual Symp, 1999 from Amer Med Informatics Assoc.
(<http://www.medicine.ucsd.edu/f99/index.>). (Broker)

Baclawski, K., R. Futrelle, N. Fridman, and M. J. Pescitelli. Database techniques for biological materials & methods. *ISMB* 1: 21-8, 1993. (Bio-Informatics)

Baker, D. D. et al. Assurance: The power behind PCASSO Security. In: *AMIA '99 (Amer Med Informatics Assoc), Annual Symposium*, 1999. (Broker)

Baumann, P. A database array algebra for spatio-temporal data and beyond. *Lecture Notes in Computer Science*, 1649: 76, 1999. (Temporal)

Baxevanis, A. D. Information retrieval from biological databases. *Methods Biochem. Ana.l* 9: 98-120, 1998. (Bio-Informatics)

Bertino, E., C. Bettini, E. Ferrari, and P. Samarati. An access control model supporting periodicity constraints and temporal reasoning. *ACM Transactions on Database Systems* 23(3): 233, 1998. (Temporal)

Bidgood, W. D., Y. Safadi, M. Tucker, F. Prior, G. Hagan, and J. E. Mattison. The role of digital imaging and communications in medicine in an evolving healthcare computing environment: the model is the message. *J. Digit. Imaging* 11(1): 1-9, 1998. (Broker)

Blake, J. A., and C. J. Bult. Biological Databases on the Internet. Publisher location: name, year? (Bio-Informatics)

Borok, L. S. Data mining: sophisticated forms of managed care modeling through artificial intelligence. *J. Health Care Financ.* 1997. (Software)

Breen, E. J., G. H. Joss, and K. L. Williams. Dynamic arrays for fast, efficient, data manipulation during image analysis: a new software tool for exploratory data analysis. *Comput. Methods Programs Biomed.* 37(2): 85-92, 1992. (Software)

Breen, E. J., and K. L. Williams. Hash function performance on different biological databases. *Comput. Methods Programs Biomed.* 28(2): 87-91, 1989. (Bio-Informatics)

Brinkley, J. F., S. W. Bradley, J. W. Sundsten, and C. Rosse. The digital anatomist information system and its use in the generation and delivery of Web-based anatomy atlases. *Comput. Biomed. Res.* 30(6): 472-503, 1997. (Broker)

Bryant, S. H., and C. W. V. Hogue. Structural Neighbors and Structural Alignments: The Science Behind ENTREZ/3D. Presented at the IUCr Macromolecular Crystallography Computing School, 1996. (Bio-Informatics)

Carhart, R. E., H. D. Cash, and J. F. Moore. StrateGene: object-oriented programming in molecular biology. *Comput. Appl. Biosci.* 4(1): 3-9, 1988. (Bio-Informatics)

Chan, S. S., and Q. Li. Developing an Object-Oriented Video Database System with Spatio-Temporal Reasoning Capabilities. *Lecture Notes in Computer Science* 1728: 47, 1999. (Temporal)

Chang, C., and B. Schatz. Efficient Information Retrieval - performance and Implications of Semantic Indexing in a Distributed Environment. *SIGIR Forum* 1, 1999. (Broker)

Chang, C. T. K., et al. Performance and Implications of Semantic Indexing in a Distributed Environment. Publisher location?: name?, 1999. (Bio-Informatics)

Chen, I. M., A. S. Kosky, V. M. Markowitz, and E. Szeto. Advanced query mechanisms for biological databases. *ISMB* 6: 43-51, 1998. (Temporal)

Cherry, J. M., S. W. Cartinhour, ACEDB: a tool for biological information. In: Adams MD, Fields, C., Venter, C. J. (eds). Automated DNA sequencing and analysis. 1994. (Bio-Informatics)

Cheung, K., P. M. Nadkarni, D. Shin. A Metadata Approach to Query Interoperation Between Molecular Biology Databases. *Bioinformatics* 14 (6). 1998. (Bio-Informatics)

Chung, S. Y., L. Wong. Kleisli: a new tool for data integration in biology. *Trends Biotechnol* 17(9):351-5. 1999 Sep. (Software)

Colosimo, A., E. Rota, P. Omodeo. A Hypercard program for the identification of biological specimens [published erratum appears in *Comput Appl Biosci* 1991 Apr;7(2):285], *Comput Appl Biosci*. 7(1):63-9. 1991 Jan. (Bio-Informatics)

Coupaye, T. Wrapping SRS with CORBA: from textual data to distributed objects. *Bioinformatics* 15(4):333-8. 1999. (Software)

Davidson, S. B., C. Overton, P. Buneman. Challenges in integrating biological data sources. *J Comput Biol*. 2(4):557-72. 1995 Winter. (Bio-Informatics)

Degrave, W., A. B. De Miranda, A. Amorim, A. Brandao, M. Aslett, Vandeyar. TcruziDB, an integrated database, and the WWW information server for the Trypanosoma cruzi genome project. *Mem Inst Oswaldo Cruz*. 92(6):805-9. 1997 Nov-Dec. (Bio-Informatics)

Djeraba, C., H. Briand. Temporal and Interactive Relations in a Multimedia Database System. Lecture notes in computer science. no. 1242. 1997. (Temporal)

Doiallo, B. A Review of Database Mgt Sys Suitable for Neuroimaging. *Meth Inform Med*. 1999. (Broker)

ECBS '00 Eng of Computer-Based Sys Conf from Institute of Electrical and Electronics Engineers (IEEE), Computer Society. (Broker)

Etzold, T., P. Argos. SRS--an indexing and retrieval tool for flat file data libraries. *Comput Appl Biosci* 9(1):49-57. 1993 Feb. (Software)

Etzion, O. Temporal Databases: Research and Practice. Lecture Notes in Computer Science; 1399. ISBN 0302-9743. 1998. (Temporal)

Gal, A., O. Etzion. A Multiagent Update Process in a Database with Temporal Data Dependencies and Schema Versioning. *IEEE transactions on knowledge and data engineering*. 10(1): 21-38. 1998. (Temporal)

Ganguly, S., M. Noordewier. PROXIMAL: a database system for the efficient retrieval of genetic information. *Comput Biol Med* 26(3):199-207. 1996 May. (Bio-Informatics)

- Garshnek, V., F. M. Burkle. Applications of Telemedicine and Telecommunications to Disaster Medicine: Historical and Future Perspectives. *J. Am. Med. Inform. Assoc.* 6: 26-37. 1999. (Broker)
- Gerneth, F., R. Haux, C. A. Muller. Data modeling for immunological and clinical data of leukemia and myasthenia patients. *Methods Inf Med* 31(2):136-46. 1992 Jun. (Med-Informatics)
- Giudicelli, V., D. Chaume, M. P. Lefranc. IMGT/LIGM-DB: a systematized approach for ImMunoGeneTics database coherence and data distribution improvement. *ISMB* 6:59-68. 1998. (Software)
- Goodman, N., S. Rozen, L. D. Stein, A. G. Smith. The LabBase system for data management in large scale biology research laboratories. *Bioinformatics* 14(7):562-74. 1998. (Bio-Informatics)
- Goralwalla, I. A., D. Szafron, O. M. Tamer, R. J. Peters. A temporal approach to managing schema evolution in object database systems. *Data & knowledge engineering.* 28(1):73. 1999. (Temporal)
- Graves, M. Data note system for capturing laboratory data. *Genomics.* 15;43(2):232-6. 1997 Jul. (Bio-Informatics)
- Grosso, W. E., H. Eriksson. Knowledge Modeling at the Millennium (The Design and Evolution of Protege-2000). 1999. [SMI-1999-0801]. (Software)
- Hearne, C., Z. Cui, S. Parsons, S. Hajnal. Prototyping a genetics deductive database. *ISMB* 2:170-8. 1994. (Bio-Informatics)
- Hripcsak, G. WebCIS: Large Scale Deployment of a Web-Based Clinical Information System. In *AMIA '99 (Amer Med Informatics Assoc) Annual Symp.* 1999. (Broker)
- Hu, J., C. Mungall, D. Nicholson and A.L. Archibald. Design and implementation of a CORBA-based genome mapping system prototype. *Bioinformatics* 14(2):112-20. 1998. (Broker)
- Jensen, C. S., C. E. Dyreson, M. Boehlen, J. Clifford. The Consensus Glossary of Temporal Database Concepts -February 1998 Version. *Computer Science.* 1399:367. 1998. (Temporal)
- Jordan, M. M. Structure and function of databases. *Biochem Soc Trans* 17(5):831-3. 1989 Oct. (Bio-Informatics)
- Joubert, M., F. Volot, D. Fieschi, M. Fieschi. Conceptual integration of information databases into an Intranet. *MEDinfo* 9(1):161-5. 1998. (Med-Informatics)

- Jun, C., J. Jie. Event-based spatio-temporal database design. *International archives of photogrammetry and remote sensing*. 32(4):105. 1998. (Temporal)
- Karp, P. D.. A strategy for database interoperation. *J Comput Biol*. 2(4):573-86. 1995 Winter. (Temporal)
- Karp, P. D., S. Paley. Integrated access to metabolic and genomic data. *J Comput Biol*. 3(1):191-212. 1996 Spring. (Bio-Informatics)
- Kitakami, H., Y. Tateno, T. Gojobori. Toward unification of taxonomy databases in a distributed computer environment. *ISMB* 2:227-35. 1994. (Software)
- Law, V., H. S. Goldberg, P. Jones, C. Safran, A component-based problem list subsystem for the HOLON testbed. Health Object Library Online. *Proc AMIA Symp* 411-5. 1998. (Broker)
- Leach, A. R., J. Bradshaw, D. V. Green, M. M. Hann, J. J. Delany 3rd. Implementation of a system for reagent selection and library enumeration, profiling, and design. *J Chem Inf Comput Sci* 39(6):1161-72. 1999 Nov-Dec. (Bio-Informatics)
- Lemaitre, D., M. C. Jaulent, U. Gunnel, A. M. Demiris, P. A. Michel, A. M. Rassinoux, B. Goransson, E. Olsson, P. Degoulet. ARTEMIS-2: an application development experiment with the HELIOS environment. *Comput Methods Programs Biomed* 45 Suppl:S127-38. 1994 Dec. (Bio-Informatics)
- Lewis, S. Design issues in developing laboratory information systems. In: Automated DNA sequencing and analysis. 1994. (Bio-Informatics)
- Liefke, H. & Davidson, S. B. Updating Complex Value Databases. 1998. (Software)
- Lindek, S., R. Fritsch, J. Machtynger, P.A. de Alarcon, M. Chagoyen. Design and realization of an on-line database for multidimensional microscopic images of biological specimens. *J Struct Biol* 125(2-3):103-11. 1999 Apr-May. (Broker)
- Lorber, M., K. Johnson, B. Kross, P. Pinsky. A Study of the Temporal Variability of Atrazine in Private Well Water. Part I: Study Design, Implementation, and Database Development. *Environmental monitoring and assessment*. 47(2) 1997. (Temporal)
- Macauley, J., H. Wang and N. Goodman. A model system for studying the integration of molecular biology databases. *Bioinformatics* ;14(7):575-82. 1998. (Bio-Informatics)
- Malet, G. et al. A Model for Enhancing Internet Medical Document Retrieval with Medical Core Metadata. *J. Am. Med. Inform. Assoc.* 6: 163-172. 1999. (Broker)

Martinez, R., J. Rozenblit, J. F. Cook, A. K. Chacko, H. L. Timboe, Virtual management of radiology examinations in the virtual radiology environment using common object request broker architecture services. *J Digit Imaging* 12(2 Suppl 1):181-5. 1999 May. (Broker)

Marx, B. Introduction to Database modeling in BioInformatics. In :The EBI Online Manual on Molecular Biology Databases, Apweiler, R., Lopez, R., Marx, B. 1999. (Bio-Informatics)

Masunaga, Y. New Generation Database Technologies for Collaborative Work Support and Spatio-Temporal Data Management. *IEICE transactions on information and systems*. 82(1): 45. 1999. (Temporal)

Mazzola, S., F. G. Mazzola, C. Di Marco. Organization and management of a database in the respiratory physiology. *Comput Biol Med*. 20(5):341-50. 1990. (Bio-Informatics)

Musen. M.. Stanford Medical Informatics: Uncommon research, common goals. *MD Computing* (1):47-50. 1999. (Temporal)

Nadkarni, P. M.. Organization of Heterogeneous Scientific Data Using the EAV/CR Representation. *J Am Med Inform Assoc*. 6:478-493. 1999. (Broker)

Nguyen, J. H.. Integration of Temporal Reasoning and Temporal Data Maintenance Into a Reusable Database Mediator to Answer Abstract, Time-Oriented Queries: The Tzolkin System. *Journ Intelligent Info Sys* 13(1/2):121-145. 1999. (Temporal)

Nigrin, D. J., K. Kohane, Scaling a Data Retrieval and Mining Application To The Enterprise Wide Level. 1999. (Software)

Nigrin, D. J., K. Kohane, Temporal Expressiveness in Querying a Time-stamp based Clinical Database. *J. Am. Med. Inform. Assoc* 7:152-163. 2000. (Temporal)

Noy, N. F., Knowledge Acquisition Interfaces for Domain Experts: An Empirical Eval of Protégé-2000. SMI*-2000-0825. (Software)

O'Connor, M. J. A Specification for a Temporal Query Systems. SMI*-1999-0816. (Temporal)

O'Connor, M., R. Coleman, M. K. Goldstein, M. Musen. (1999). Integrating a Modern Knowledge-Based System Architecture with a Legacy VA Database: The ATHENA and EON Projects at Stanford. *AMIA '99*, Washington, D.C. SMI-1999-0795. 1999. (Temporal)

Ohe, K. A hospital information system based on Common Object Request Broker Architecture (CORBA) for exchanging distributed medical objects--an approach to future environment of sharing healthcare information. *MEDinfo* 9(2):962-4. 1998. (Broker)

Ozsoyoglu, G., R. T. Snodgrass. Temporal and Real-Time Databases: A survey. *IEEE Transactions On Knowledge and Data Engineering*, 1995. (Temporal)

Provost, F. J., D. N. Hennessy. Distributed machine learning: scaling up with coarse-grained parallelism. *ISMB* 2:340-7. 1994. (Broker)

Rodriguez, L., H. Ogata, Y. Yano. An Access Mechanism for a Temporal Versioned Object-Oriented Database. *IEICE transactions on information and systems*. 82(1): 128. 1999. (Temporal)

Russell, K., S. Moore. Research integration. *Breast Cancer Res Treat* 31(2-3):377-80. 1994. (Med-Informatics)

Sakamoto, N., S. Goto, T. Takagi. A deductive database system for analyzing human nucleotide sequence data. *Int J Biomed Comput* 36(3):171-9. 1994 Jul. (Temporal)

Saldanha, J., J. R. Eccles. GENPRO: automatic generation of Prolog clause files for knowledge-based systems in the biomedical sciences. *Comput Methods Programs Biomed*. 28(3):207-14. 1989 Mar. (Broker)

Samarabandu, J. K., R. Acharya, P.C. Cheng. Visualization and interactive exploration of multidimensional confocal images. *Comput Med Imaging Graph*. 17(3):183-8. 1993 May-Jun. (Software)

Satou, K., E. Furuichi. A Deductive Database System PACADE fir analyzing 3-D and Secondary Structures of Protein. *Computer Applications in the Biosciences*. 9(3):259-265. 1993. (Biology)

Scherrer, J. R. Health Care Information System Architecture (HISA) and its Middleware Models. In AMIA '99 (Amer Med Informatics Assoc) *Annual Symp*. 1999. (Broker)

Sellis, T. Research Issues in Spatio-Temporal Database Systems. Lecture notes in *computer science*. 1651: 5. 1999. (Temporal)

Shahar, Y., M. Musen. Eval of Temporal-Abstraction Knowledge-Acquisition Tool. (SMI*-1999-0808). 1999. (Temporal)

Shahar, Y. Semi-Automated Entry of Clinical Temporal-Abstraction Knowledge. *JAMIA* 6(6):494-511. 1999. (Temporal)

Shankar, R. D. Justification of Automated Decision-Making: Medical Explanation or Medical Argument (SMI*-1999-0794). 1999. (Broker)

Silberzahn, N., F. Lefevre, P. Le Beux, A new Object Request Broker architecture in implementing a medical picture retrieval system. *Stud Health Technol Inform.* 43(B):586-90. 1997. (Broker)

Singh, G. B. Detecting biological patterns. The integration of databases, models, and algorithms. *Methods Mol Biol.* 132:431-45. 2000. (Bio-Informatics)

Snodgrass, R. Special Series on Temporal Database (Part 1). *Database programming & design.* 11(6): 46. 1998. (Temporal)

Snodgrass, R. Special Series on Temporal Database (Part 2): Querying Valid-Time State Tables. *Database programming & design.* 11(7): 60. 1998. (Temporal)

Snodgrass, R. Special Series on Temporal Database (Part 3): Modifying Valid-Time State Tables. *Database programming & design.* 11(8): 72 . 1998. (Temporal)

Snodgrass, R. Special Series on Temporal Database (Part 5): Temporal Support in Standard SQL. *Database programming & design.* 11(10): 44. 1998. (Temporal)

Sorace, J. M., K. Canfield. Collaborative bioinformatics: data warehouses for targeted experimental results. *J Interferon Cytokine Res.* 18(9):799-802. 1998 Sep. (Bio-Informatics)

Springer, G. K., T. B. Patrick. A service-oriented information sources database for the biological sciences. *ISMB* 1:380-6. 1993. (Broker)

Stead, W. W., Integration and Beyond: Linking Information from Disparate Sources and into Workflow. *J Am Med Inform Assoc.* 7:135-145. 2000. (Broker)

Stein, L. D., S. Cartinhour, D. Thierry-Mieg, J. Thierry-Mieg, JADE: an approach for interconnecting bioinformatics databases. *Gene.* 16;209(1-2):GC39-GC43. 1998 Mar. (Bio-Informatics)

Stein, L. D., J. Thierry-Mieg. Scriptable access to the *Caenorhabditis elegans* genome sequence and other ACEDB databases. *Genome Res.* 8(12):1308-15. 1998 Dec. (Bio-Informatics)

Tange, H. J., A. Hasman. Medical Narratives in Electronic Medical Records. *International Journal of Medical Informatics* 46:7-29. 1997. (Med-Informatics)

Taylor, S. A, I. D. Chivers. The assessment of biomarkers to detect nephrotoxicity using an integrated database. *Environ Res.* 75(1):23-33. 1997 Oct. (Bio-Informatics)

Temporal Aggregation in Active Database Rules. *SIGMOD* 26(2): 440-52. 1997. (Temporal)

Van de Velde, R. Framework for a clinical information system [In Process Citation]. *Int J Med Inf.* 57(1):57-72. 2000 Jan. (Broker)

Walsh, S., M. Anderson. ACEDB: A Database for Genome Research. *Methods Biochem Anal.* 1998. (Bio-Informatics)

Wang, C. O., K. S. Kaihara. Dynamic link between ECG and clinical data by a CORBA-based query engine and temporal mapping. *Proceedings of the AMIA Symposium.* 1997. (Broker)

Wang, C. O., K. S. Kaihara, Dynamic link between ECG and clinical data by a CORBA-based query engine and temporal mapping. *Proc AMIA Annu Fall Symp* 27-31. 1997. (Broker)

Wang, X. S., C. Bettini, A. Brodsky, S. Jajodia. Logical Design for Temporal Databases with Multiple Granularities. *ACM transactions on database systems.* 22(2), 1997. (Temporal)

Weinshenker, B. G. Database in MS Research: Pitfalls and Promises. *Multiple Sclerosis.* 1999. (Med-Informatics)

Wijzen, J. Temporal FDs on Complex Objects. *ACM transactions on database systems.* 24, no. 1:127. 1999. (Temporal)

Winslett, M. New database technology for nontraditional Applications. 1993. (Bio-Informatics)

Wu, Y., S. Jajodia, X. S. Wang. Temporal Database Bibliography Update. Lecture notes in *computer science.* no. 1399. 1998. (Temporal)

Yearworth, M., S. Battle. Workflow management for multimedia information in clinical laboratories. *Comput Methods Programs Biomed* 55(1):1-9. 1998 Jan. (Broker)

Yoshida, K. Maintenance and control of BIAS system which consists of a bio-instrumental LAN including a mini-computer. *Tanpakushitsu Kakusan Koso.* 39(11):2038-50. 1994 Aug. (Broker)