Knowledge Mining Application in ISHM Testbed

William J. McDermott	Peter Robinson	Daniel P. Duncavage
650-604-5650	650-604-3513	281-792-5478
B.McDermott@nasa.gov	robinson@mail.arc.nasa.gov	Daniel P. Duncavage@nasa.gov
David A. Maluf	SAIC, NASA Ames Research Center	NASA Johnson Space Center
650-604--0611	Moffett Field, CA 94035-1000	Houston, TX 77058--396
maluf@email.arc.nasa.gov	Mohana Gurram	Kevin Bass
Rick Alena	650 604-4683	281- 792-5496
650-604-0262	mgurram@email.arc.nasa.gov	kbass@mail.arc.nasa.gov
ralena@mail.arc.nasa.gov	Universities Space Research Association, NASA Ames Research Center	QSS Group, Inc, NASA Johnson Space Center
NASA Ames Research Center	Moffett Field, CA 94035-1000	Houston, TX 77058--396
Moffett Field, CA 94035-1000		

Abstract—The scope of Integrated System Health Management (ISHM) is the end-to-end autonomous management of subsystems, systems and systems of systems. Among ISHM's components is management of data, information, and knowledge to detect and diagnose anomalies, and propose mitigation procedures for crew implementation. Complex systems such as the International Space Station (ISS) generate a large amount of data, over 160,000 documents that are distributed across disparate systems. But this data is fragmented and applications are unable to interoperate. In addition, the lack of a seamless mechanism for ensuring use of latest, best information hinders development of much needed modeling, simulation and visualization systems that could increase the operating safety of a complex system.

To reduce development time for ISHM systems, builders of advanced diagnostic systems require rapid retrieval of precise information from various sources spread across multiple, distributed, heterogeneous databases.[1]

To meet this challenge, a technological leap that goes beyond the bounds of today's understanding of documents is needed. Software developed by NASA Ames researchers in the Advanced Knowledge Exploration Networks (AKEN) Lab that enables diagnostic systems and distributed information sources to be linked into an Information Grid is one such technology. Funded through the Integrated Testbed Prototypes (ITP) Project, in 2005 AKEN installed Netmark's Context Based Retrieval system at Johnson Space Center to incorporate this technology into advanced diagnostic systems. For the ITP Project this technology was given the name Knowledge Mining Application (KMA). KMA was responsible for integrating documents from a range of information sources so that an advanced caution and warning diagnostic system could retrieve precise information on ISS events in question.

TABLE OF CONTENTS

1. INTRODUCTION

The Challenge—The challenge is seamless integrated access to multiple, distributed, and heterogeneous information sources

The Problem—Complex systems such as ISS and Crew Exploration Vehicle (CEV) generate huge amounts of reports, procedures, schematic diagrams, drawings and other documents that are distributed across disparate systems. This highly distributed information must be securely accessible and usable across the Enterprise. However, humans accessing information about ISS and CEV face many challenges. To find the best information on an item of interest engineers must query a series of databases, using a different identifier each time. When something doesn't match the engineer has to sort through documents to make the missing connection. Integrating information sources spread across distributed, heterogeneous data bases is a

[1]U.S. Government work not protected by U.S. copyright

[1]IEEEAC 1296, Version 3, Updated Dec, 5 2005

challenge for NASA. Middleware and data-integration systems that address this problem have been around for some time [1]. However, current technology requires investing in complex middleware and expensive schemas to create a large application of many requirements. Each application requires schemas or views for each source, and mapping between schemas or creating global views to perform the integration [2]. With this approach, application integration can be very expensive. AKEN experience developing data integration systems for NASA Enterprise has led us to an approach that is scalable, flexible, and cost-effective.

Our Approach—Input for the vast majority of integration applications at NASA is made up of documents, spreadsheets, reports and presentations[2]. Also, requirements for query center around extracting select sections from documents and using them to create new documents or performing keyword searches on documents. Our approach is to develop data management and integration systems optimized for these common requirements. With our approach, formal definition of schemas is kept to a minimum. Complex middleware for integration is not required. Also, specific integration capabilities can be readily defined, dynamically by the application because functionality for integration is delegated to the client [3].

Terms and Definitions—Data, Engineering Data, Information and Knowledge are defined as follows:

Data − Comprise facts, observations, or perceptions. Alone, data represent raw numbers or assertions, and may therefore be devoid of context, meaning or intent [5].

Engineering Data − Data that has been given an offset or amplification; data with engineering units [6].

Information − Includes data that possess context, relevance, and purpose. Information typically involves the manipulation of raw data to obtain a more meaningful indication of trends or patterns in the data [1]. Information pertains to known facts that result from processed data, and inferred facts from data by using knowledge and other facts [5].

Knowledge − Knowledge in an area is defined as justified beliefs about relationships among concepts relevant to that particular area. Intrinsically different from information knowledge consists of facts and inference rules used for reasoning. Knowledge can be procedural or declarative [6].

ITP and Knowledge Management

Relevance to NASA—NASA's future exploration systems programs will require advances in ISHM. Furthermore, requirements for modeling, simulation and visualization that support ISHM are driving the need for new technologies in NASA's Exploration Systems Research and Technology (ESR&T) Formulation Plan. Funded through ESR&T, the Integrated Testbed Prototypes (ITP) Project is one of the projects addressing these requirements. The ITP Project is developing technologies to complete fully functioning ISHM testbeds and prototype implementations. The testbeds will be used to develop and validate technology needed for monitoring and managing diverse subsystems/systems, particularly the systems-of-systems that will be required to implement NASA's Vision for Exploration.

ITP is an Information System—Among ISHM's components are 1) Data, information and knowledge handling and interoperability standards, and, 2) A system for managing data, information, and knowledge to detect and diagnose anomalies, and propose mitigation procedures [4].

ITP Framework—The ITP project is developing an infrastructure that is persistent and portable to new programs, providing the ability to continually develop and test ISHM technologies. An important product of ITP is implementation of a testbed at Johnson Space Center for the International Space Station. The purpose of this testbed implementation is to rigorously test technologies, such as diagnostic models, using a real NASA system. In 2005, a prototype ISHM implemented and validated (independent of regular operations) on an ISS subsystem was begun. The prototype testbed implementation consists of ITP framework and software for diagnosis, prognosis and remediation of system anomalies and knowledge mining and management. The ITP infrastructure/framework is an open standards-based, network-centric architecture and component set designed to allow maximum flexibility for solutions to be developed and tested without the cost of tight integration. The Knowledge Mining Application (KMA) and Collaborative Environment (CE), is a component of the ITP framework, shown in figure 1.1.

[2] We use the term 'documents' to include documents in formats such as Word, PDF, HTML, XML or others, spreadsheets, presentations, reports, etc., henceforth

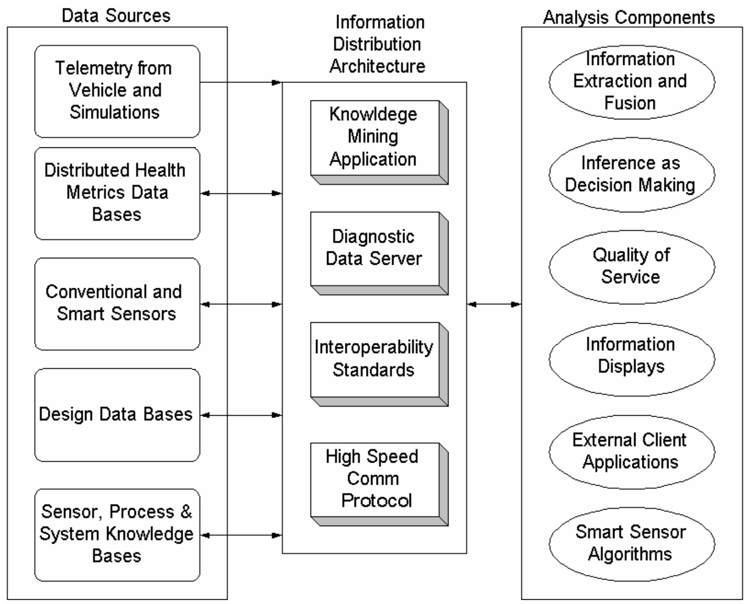

Figure 1.1 ITP Framework and KMA's relationship to Client Applications

KMA Goals, Objectives, and Focus—Our goal is to reduce development time for advanced diagnostic systems running in an ISHM system. Our objective is to create knowledge management capabilities that integrate diagnostics with flight rules and procedures to aid rapid recovery and to help determine mission consequences of failures. Our focus is guided by the fact that researchers who are using an ISHM testbed to develop advanced diagnostic models will need documents, human readable data – specifications, schematics, drawings, XML. KMA integrates data, information, and knowledge by providing these domain experts with tools and a collaborative environment for mining and sharing knowledge from document repositories of the complex system in question. KMA accomplishes this through the use of NETMARK, a contextual search and information retrieval tool developed at NASA Ames Research Center [3]. In FY05 Netmark was installed at the Johnson Space Center and matured to meet KMA functionality requirements of ITP. The KMA was assessed using a complex, real environment (ISS), which entails crossing multiple organizations and technical standards

2. COURSE OF ACTION OBJECTIVE

KMA and NX Knowledge Netmark

A partnership between NASA Ames Research Center and Xerox has resulted in a new system called NX- Knowledge Network (NX). NX incorporates NASA Ames' NETMARK contextual search software and content management with collaboration software from Xerox's global research centers. NX is already used to support collaborative research across the various missions and project teams at NASA Ames and provided a readily available, economical collaborative environment (CE) for the ITP project. The relationship between KMA and Client Applications within the ITP framework is shown in figure 2.3.

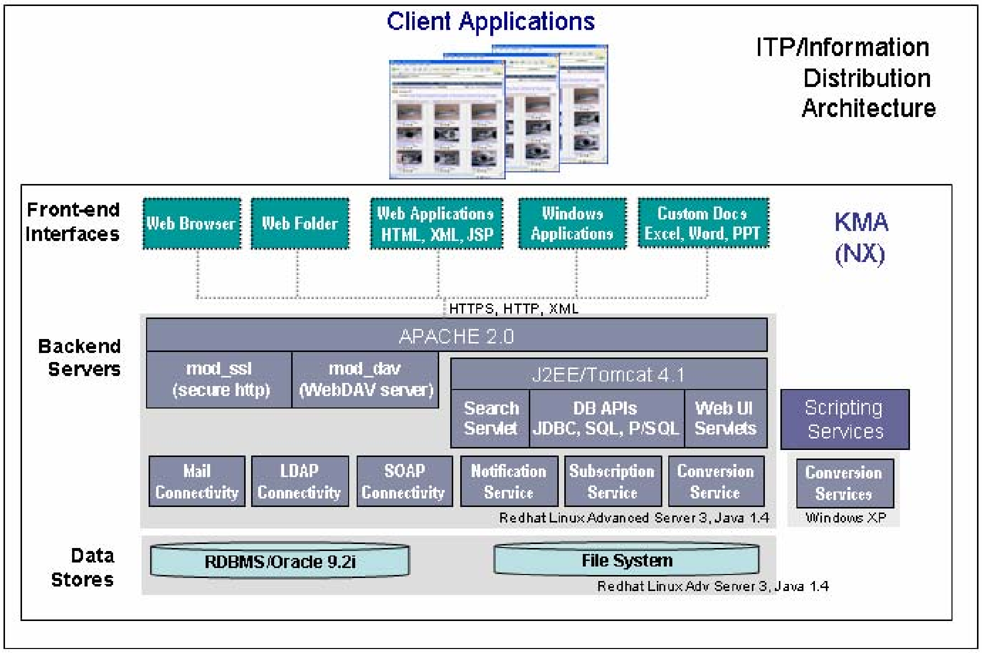

Figure 2.3 KMA/NX Architecture

NX Overview—NX is a platform that supports applications in a variety of scientific fields. NX's purpose is to control and interoperate with every block in a document, email, spreadsheet, presentation, database across the lifecycle.

Using NX technology, KMA creates an Information Grid that will provide seamless integration of distributed heterogeneous information resources for distributed scientific and engineering applications. A novel architecture that enables the creation of such an Information Grid, the NX is built upon three standards from the World Wide Web Consortium Architecture Domain and the Internet Engineering Task Force:

(1) HTTP: Hypertext Transfer Protocol

(2) XML – a five-year old standard that defines a syntax for exchange of logically structured information

(3) Web DAV – a widely supported four-year old standard for distributed management of web resources.

NX/KMA leverages WebDAV for management of arbitrary information resources including information processing services. Through a combination of these international protocols, universal database record identifiers, and physical address data types, KMA provides a number of capabilities for managing distributed and heterogeneous information resources such as the following:

(1) Storing and retrieving information about resources using properties

(2) Locking and unlocking resources to provide serialized access

(3) Getting and putting information in heterogeneous formats

(4) Copying/moving/organizing resources through hierarchy and network relations

(5) Automatic decomposition of information into a query-able XML database

(6) Context+content querying of information in the XML database

(7) Sequencing workflows of information processing tasks

(8) Seamless access to information in diverse formats and structures

The operation of KMA's WebDAV HTTP API is shown in figure 2.4

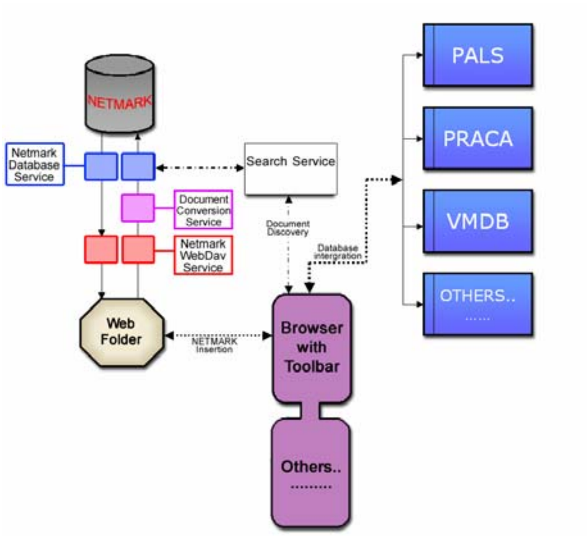

Figure 2.4 KMA HTTP API – Doc. Insertion, Discovery + Integrating Multiple ISS Data Bases

KMA thus represents a flexible, high-throughput open architecture for managing, storing, and searching unstructured or semi-structured data. KMA provides automatic data management, storage, retrieval, and discovery in transforming large quantities of highly complex and constantly changing heterogeneous data formats into a well-structured, common standard.

Context+Content Search—KMA takes advantage of the boundaries that demarcate the location of information within a document. KMA's Context Based Retrieval mechanism is illustrated in figure 2.5.

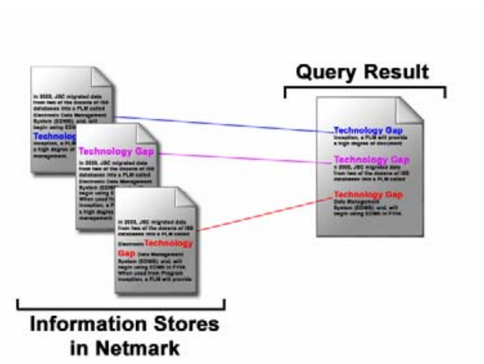

Figure 2.5 Context-Based Retrieval of relevant text fragments and documents

A section heading, such as "Technology Gap", that appears before a paragraph can throw light on the meaning of text within. Aware of the meaning conveyed by section-headings, KMA uses this information as the context for the query, thus enabling KMA to return only relevant documents.

KMA-Client Application Integration

ISS Problem Context—The ISS flight control team uses a diagnostic process to understand events and keep their effects from spreading. The description of a caution and warning event that the flight control team receives is usually driven by an event code. Display of descriptions of events triggers the flight control team's efforts to develop an appropriate response. The perspective of Flight Operations is shown in Figure 2.6.

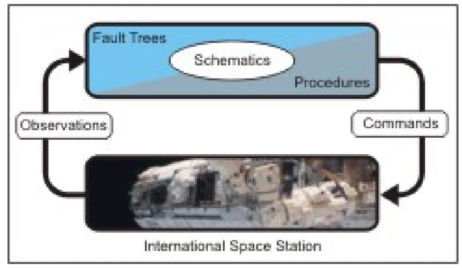

Figure 2.6 ISS Control Methodology

The information on the display may only mean that an event appears to have happened, it does not show definitively what has happened. The flight control team must understand what has happened and what they can do to handle the event. To reach this understanding, the flight control team consults readings from other sensors, fault trees, schematics and procedures, each information source possibly pointing to others. The flight control team follows this chain of information to procedures that define corrective action. Then astronauts and flight controllers issue commands to put changes into effect. For example, to keep a problem from spreading, an astronaut or flight controller would command ISS to move a pump module off-line and reconfigure a backup pump.

ISStrider Project—The ISStrider project is developing model-based reasoning, visualization and document retrieval capabilities for mission and spacecraft operations [7]. These include the Caution and Warning Fusion (CWF) capability for root-cause determination, the Caution and Warning Cube (CW^3) for visualization and the Real-time Knowledge Management (RKM) capability for document retrieval. ISStrider Project is applying these capabilities to the ISS domain. The lessons from the ISS legacy spacecraft will allow NASA to develop approaches applicable to future spacecraft and mission operations concepts.

ISS Document Set—KMA-ISStrider integration involved bringing together data and documents about the following:

(1) ISS Emergency Caution Warning Advisory History

(2) Command Control System

(3) Navigational Control System

(4) Station Operations Data File (SODF) [7]

(5) Caution and Warning Tables

ISStrider scenario − ISStrider performs the following steps:

(1) Replay Caution and Warning (CW) Events for 2002 from History files

(2) Select CW Events 5014 and 5392

(3) Fetch Station Operation Data Files associated with a CW event

(4) Analyze root cause of CW events 5014 and 5392

KMA provides Real Time Knowledge Mining to ISStrider—KMA traces from an event code on Command and Data Handling, a major ISS subsystem, to a crew response. A Crew response is described in a list of steps found in SODF procedures. KMA passed URLS of the relevant procedures back to ISStrider. The relationship between KMA and Advanced Caution and Warning/diagnostic applications such as ISStrider is shown in figure 2.2.

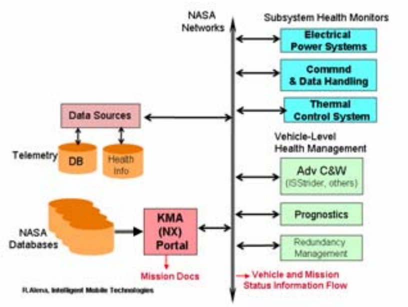

Figure 2.2

Conversion and loading of ISS Document Set into Netmark Data Base—For the demonstration, four MSWord documents and one Excel spreadsheet containing event code definitions were converted to XML and stored in NX's data base.

ISS Standard Event-Codes—The following were retrieved from the ISS procedures site:

(1) Event-code definitions [8]

(2) Advisory-code definitions [9]

(3) Caution-code definitions [10]

(4) Warning-code definitions [11]

(5) Alert, Caution, Warning tables contain messages about ISS subsystems

Sample Event Code Table Entry—A message about an event is made up of data values across a row in a table with the following column headings: System, Message Text, Crew Response, Event Code, Must Respond < 1 Orbit. Data in each message and the elements within the message are converted into XML. An example of a message for event code 5043 from the Communication and Control System (CCS) Caution Table is shown in Table 2.1.

Table 2.1 Example of Messages in CCS Caution Table

System	Message Text	Crew Response	Event Code	Respond < 1 Orbit
Communications and Tracking System				
CNT	UHF power supply current out of limit	Suppressed. **MCC** using message as system anomaly indication. **MCC** will execute procedure {2.702 UHF 1 ORU DEACTIVATION} (SODF: C&T: NOMINAL: UHF), then {2.703 UHF 2 ORU ACTIVATION}, steps 1 to 15 (SODF: C&T: NOMINAL: UHF), OR {2.704 UHF 2 ORU DEACTIVATION} (SODF: C&T: NOMINAL: UHF), then {2.701 UHF 1 ORU ACTIVATION}, steps 1 to 15 (SODF: C&T: NOMINAL: UHF).	5043	

KMA-Client Application Communication—The client application tells KMA to retrieve relevant data and documents by invoking the NXProcessor, a Practical Extraction and Reporting Language (PERL)/Common Gateway Interface (CGI) script. The NXProcessor executes a user-customizable configuration file that set the query's scope, search criteria, and other parameters.

Integration Scrip − KMA's general-purpose scripting capability abstracts a chain of queries to local and remote information sources into a single HTTP call. The NXProcessor searches for configuration scripts at a specific location in NX defined by the parameter "nxpath" provided when the user invokes the CGI script. The HTTP for invoking a configuration script is provided below:

<hostname>:port-number/cgi-bin/nxprocessor?nxpath=<complete URL of script to be executed>&context1=<section or element enclosing the content>content1=<content of query>>&context2=<parent section of context1>

Script Execution—Control Flow of NX configuration script that maps from event code to procedures' URLs is shown in figure 2.5. In response to the HTTP call, the script finds the event-code's classification by looking up the event code in the event-definition table. The event-definition tells the script the category of the event: Advisory, Caution, Emergency, Pressure or Warning. The script substitutes the event's classification into the scope of the query that follows. The event's classification directs the query to the table that may contain an entry for the event. The script now has enough information to query the relevant event-table. The query of the event-table returns an XML file containing procedure titles associated with the event-code in question. The titles that the script found constitute the crew response to the event. The titles also make up the criteria for a series of queries to find each procedure's URL in NX. For each procedure title, the script retrieves the URL for a PDF file with a matching procedure-title. A series of queries of procedure titles results in an XML file containing URLS. Before returning these URLS, the NX script stores an XML file in NX to record execution of the script. This metadata file lists the results of each step in the script. After inserting the metadata file the script returns to the caller XML containing URLS for the event code. ISStrider then uses NX to retrieve the precise, relevant SODFs from NX document repository. An example of URLs returned to ISStrider by the scripting process is shown in Figure 2.6.

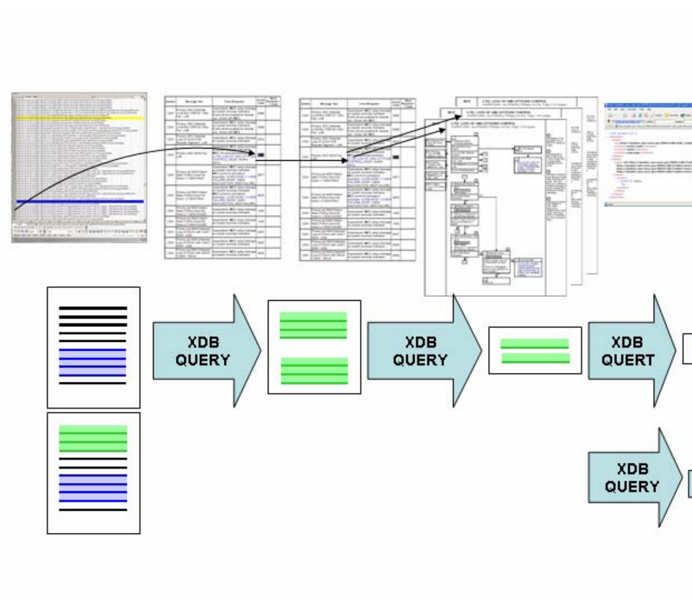

Figure 2.5 Flow of Script -- mapping from event code to ISS procedures' URLs

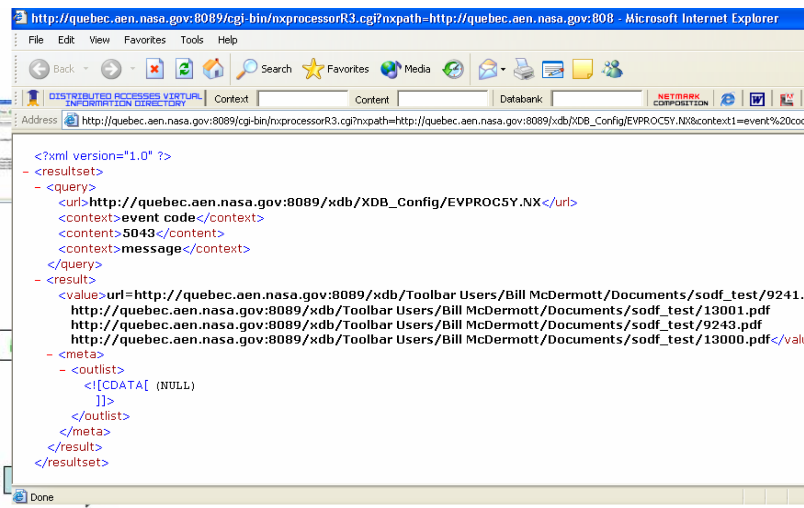

Figure 2.6 Results returned by KMA to Client Application

Related and Future Work

Implement script processing in more sustainable technology—The PERL implementation of NXprocessor is a prototype. As we capture additional requirements we expect to redesign script processing in more flexible standards-based technology such as XSLT [13].

Expand the range of information sources—We plan to expand the scope of document repositories covered by queries to include fault-trees, schematics, and software specifications from multiple ISS information sources.

Ensure use of latest, best available documents, improving quality of information—For demonstration of the KMA/Client Application integration documents were stored locally, and, KMA queried only NX's local data base. In the next phase we will store static documents locally, and retrieve more volatile documents from primary information sources, in the processing ensuring that documents are up to date.

Bridge databases by extending KMA's knowledge of ISS documents—We expect to enhance NX's context-based retrieval capabilities by adding aliases, stemming, and plurals. For example, we will produce a concordance that lists, for each SODF, all the referenced commands and their Signal Program Unique Identifiers (SPUI) and operational names. Bridging sustaining-engineering and operations domains increases the relevance of search results and reduces the current labor intensive and error prone lookup process.

Deliver the Current Working System—We will put an operational system in the hands of flight controllers and others who have a relevant need. Their experience and feedback will give us insight into KMA/ISStrider's usefulness; and, user input will guide us in future releases.

3. CONCLUSION

KMA provides data integration capabilities that reduce the time and effort to develop Client Applications. KMA accomplishes this by providing workflow that encapsulates a cascade of steps that query and extract relevant data from documents, including references to entire documents. This general-purpose scripting capability abstracts a chain of queries to local and remote information sources into a single HTTP call. A single HTTP call retrieves precise, relevant information on an event in question. Without writing code for their applications (to search for relevant documents) researchers can use NX's scripting tool to encapsulated new scripts. These scripts integrate data from different documents stored both locally and in ISS data warehouse and provide the information client-applications want.

4. REFERENCES

1. Halevy, A., Data Integration: A Status Report (Invited Talk). *German Database Conference (BTW)*, 2003 .

2. Papakonstantinou, Y. and Vassalos, V., Architecture and Implementation of an XQuery-based Information Integration Platform, *Data Engineering Bulletin* (2002).

3. Maluf, David A. Bell, David, G. Ashish., Naveen. Lean Middleware. SIGMOD 2005

4. ISHM Testbed and Prototypes Project Using the Space Station and Rocket Engine Test Stand for Exploration Technology Maturation. Project Management Plan. Advanced Space Platforms and Systems Technology Element Technology Maturation Program Exploration Systems Research and Technology.

5. Becerra-Fernandez, Irma, Avelino Gonzalez and Sabherwal, Rajiv.Knowledge Management: Challenges, Solutions and Technologies. Prentice Hall, 2004,

6. Chen, Zhengxin. Computational Intelligence for Decision Support CRC Press, 2000. ISBN 0-8493-1799-1)

7. Robinson, P., Shirley, M., Fletcher, D., Alena, R., Duncavage, D., Lee, C. Applying Model-Based Reasoning to the FDIR of the Command & Data Handling Subsystem of the International Space Station. iSAIRAS 2003.

8. Systems Operations Data File (SODF): http://mod.jsc.nasa.gov/do3/sodf/

9. CCS_ECW_04_09 (Event code définitions tables)

10. 1.101 CCS ADVISORY TABLE – MASTER (ADVISORY GND\LF1-12A\FIN 2)

11. 1.101 CCS CAUTION TABLE (CAUTION/E8 - LF1/FIN 4/MULTI)

12. 1.101 CCS WARNING TABLE (WARN/E8 - ULF1/FIN 3/MULTI)

13. XSLT, http://www.w3.org/TR/xslt.

5. BIOGRAPHIES

William J. McDermott leads the Knowldege Mining Application Element of the Integrated Testbed Protypes Project. He was a Principle Investigator of a Virtual National Airspace Simulation project. Also in Aeronautics, he was a software developer on Build Team for the Surface Movement Advisor (SMA) at Atlanta Hartsfield International Airport.

David A. Maluf leads the NASA Advanced Exploration Network laboratory (AEN), a laboratory consisting of 20+

staff with an average of 12 projects/year. He has over 70 technical publications in journals and conference proceedings, over hundreds of presentations at international conferences and is an inventor of numerous patents. He has taught courses on system engineering and databases, and has written two books. He received his PhD from McGill University and conducted post-doctoral research in information integration at Stanford University.