

Collaborative Management Environment

Thomas E. Potok, Mark T. Elmore, and Nenad Ivezic
Collaborative Technologies Research Center
Computer Science and Mathematics Division
Oak Ridge National Laboratory
Oak Ridge, Tennessee, 37831, USA

Abstract

This paper discusses an innovative use of XML as a cost-efficient alternative for acquiring, storing, querying, and publishing heterogeneous and distributed enterprise information. The web-based Collaborative Management Environment (CME) is presented in the context of a pilot system designed to automate and enhance the management of research proposal information from multiple independent research organizations within the Department of Energy (DOE).

The initial results of the system usage are very positive. To date, information may be successfully acquired, integrated, searched, and displayed using the pilot system. While the XML-based system has obvious limitations, such as limited scalability and the rapid evolution of XML technologies, the system fulfills an important role as a cost-efficient enterprise information system alternative.

Two key challenges presented themselves in the development of the pilot system. First, due to the distributed, heterogeneous, and sensitive nature of the proposal data, it is preferred that control of the data remain with the originating laboratories. Second, due to the lack of motivation for the laboratories to invest in the construction and

maintenance of traditional enterprise information systems, it is necessary to investigate a low-cost, distributed information management system.

Distributed databases, object request brokers, and various middleware packages deal very nicely with distributed, heterogeneous data, however, at quite a large cost to the owners of the data. Our approach is to use XML as an information repository that can be used to search for and present information. The produced XML files can be seen as a “database” of proposal information. We developed converters to generate raw database reports into XML files, various means of querying this information, and various ways of presenting this information, for example XSL style sheets. This approach provides a successful low-cost means of representing distributed and structured information.

Introduction

The Collaborative Technologies Research Center within the Oak Ridge National Laboratory has completed the development of a pilot system called the Collaborative Management Environment (CME). CME is a research project funded by the Department of Energy (DOE) to investigate advanced information technologies for improved management of research information across the DOE complex of national laboratories. As DOE funds a vast amount of energy-related research in a very broad range of areas, it is not surprising that each laboratory follows independent research management processes tailored to the expertise of that laboratory. Information resulting from these management processes, however, is in different formats and at different levels of granularity that, in

The submitted manuscript has been authored by a contractor of the U.S. Government under contract No. DE-AC05-96OR22464. Accordingly, the U.S. Government retains a non-exclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government Purposes.

turn, makes the overall management of DOE-funded research a difficult challenge.

The process for requesting research funding from DOE involves the submission of a research proposal, often called a Field Work Proposal (FWP). This proposal describes, among other things, why the research is significant, how the scientist(s) will use research funds, and who would benefit from the technology advance. Once a year, researchers submit proposals to DOE in a paper format requesting research funds. It is the transition from this paper-based proposal management process to an electronic-based management system that poses a significant challenge to CME.

This paper is organized as follows. First, we provide a brief account of the system requirements and related work in managing distributed data. Next, we describe our approach and the system functionality to achieve a low-cost distributed data management solution using XML as the foundational enabling technology. Then, we present our findings that follow from an application of our approach. Further, we discuss the relative merits and key issues of our approach in the light of these findings. Finally, we summarize the principal points of our paper.

Requirements

In order for the CME system to be successful, it must meet the user requirements, while also being cost efficient to the participating laboratories. The CME pilot system must have the capability to store proposal information in such a way that it can be queried based on general keywords, and by field-specific information. For example, a program manager may want to perform a general keyword search on “Neutron Physics” to investigate the research contribution proposed in this area. Or, he or she may want to find all of the proposals that have “Neutron Physics” in the title, were submitted by a principal investigator with the last name of “Smith,” all within the years 1994-1997. Likewise, the program manager may require reports and graphs derived from this information, i.e., how much money was spent by “Smith” on proposals that include the phrase “Neutron Physics” in the title. Lastly, the proposal

information must be easily viewable from a variety of platforms. Since each lab has a different presentation form for the proposal information, the program manager needs to be able to view the information in a form that closely represents the original.

Two key challenges presented themselves in the development of the pilot system. First, due to the distributed, heterogeneous, and sensitive nature of the proposal data, it is preferred that control of the data remain with the originating laboratories. Second, due to the lack of motivation for the laboratories to invest in the construction and maintenance of traditional enterprise information systems, it is necessary to investigate a low-cost, distributed information management system.

In this paper we present our pioneering work for using the eXtensible Markup Language (XML) as a cost-efficient alternative for acquiring, storing, querying, and publishing heterogeneous and distributed enterprise information. We show that XML can provide a basis for a distributed data management system alternative with significant presentation and query capabilities.

Related Work

There exists a variety of ways of addressing the issue of managing heterogeneous data, such as data warehousing, object-request brokers, and middleware tools. Data warehouses provide a means for publishing and accessing a broad range of distributed, heterogeneous data. Object-request brokers (ORBs) provide a way of accessing distributed objects, as if the objects were local to the user’s environment. Middleware tools provide a layer above the data layer yielding easy access to the distributed data. However, these approaches all assume that the owner of the data is motivated to develop and maintain a relatively costly information management system. If a user has little interest in spending significant resources to implement such a system, the above alternatives are of little value. The issue remains, how to distributed data in an environment where the data owners are not motivated to distribute their data.

Clearly, the simplest means of distributing data in today’s environment is the Web. The natural means of distributing information in this manner would be to present proposals in HTML.

The main drawback of using this approach to distributed data is the inability of HTML to efficiently store data in a structured manner. A proposal can be searched for and presented, but the functionality of searching for proposals based on key fields is missing. The eXtended Markup Language (XML) provides a way to structure proposal information so that field level searches are possible. XML was developed as a subset of the Standard Generalized Markup Language (SGML) and was recommended by World Wide Web Consortium (W3C) as a standard in February of 1998 [Bray et al. 1998]. Though XML itself has been standardized, many exterior features that would improve the usefulness of XML are still in the relatively early stages of development.

For the most part, XML is viewed as an enhancement to HTML. However, work is currently being developed that suggests that XML may provide a reasonable level of support for data publishing, exchange, and integration. For example, electronic catalogs such as the ones described in [Singh 1998] and [Lincke et al. 1998] could be integrated and published using XML technologies. Likewise, Bosak describes a model that uses XML as a 'hub' language for communication within the same industries [Bosak 1997].

We believe and have demonstrated that XML can be used as a storage, retrieval, and presentation vehicle beyond what has been previously proposed. Significantly, this application of XML can be achieved at very low cost to the data owners, while providing a basic level of data query capability [Potok et al. (1998)].

Approach

In this section we describe the goal of the CME project. Ideally, we would like to be able to capture all data associated with a Field Work Proposal. In order to do so, every field within a research proposal must be wrapped with the appropriate XML tags. For example, the title of an FWP would be surrounded by <TITLE> and </TITLE> tags. Having the information in an XML format would enable structured queries to be performed over the XML files. Furthermore, the

XML files could be located at the individual laboratories. This could provide the basis for truly a distributed, low-cost information management system.

CME is also designed to allow Principal Investigators (PIs), the researchers proposing the work, to control the presentation of their FWP information. An FWP can be thought of as a "sales brochure" where a PI explains why proposed research is valuable and worthy of funding. The PI may wish to exert individual control over the appearance of an FWP, and may wish to enhance the appearance with web technologies, such as graphs, links to related papers, audio or video clips.

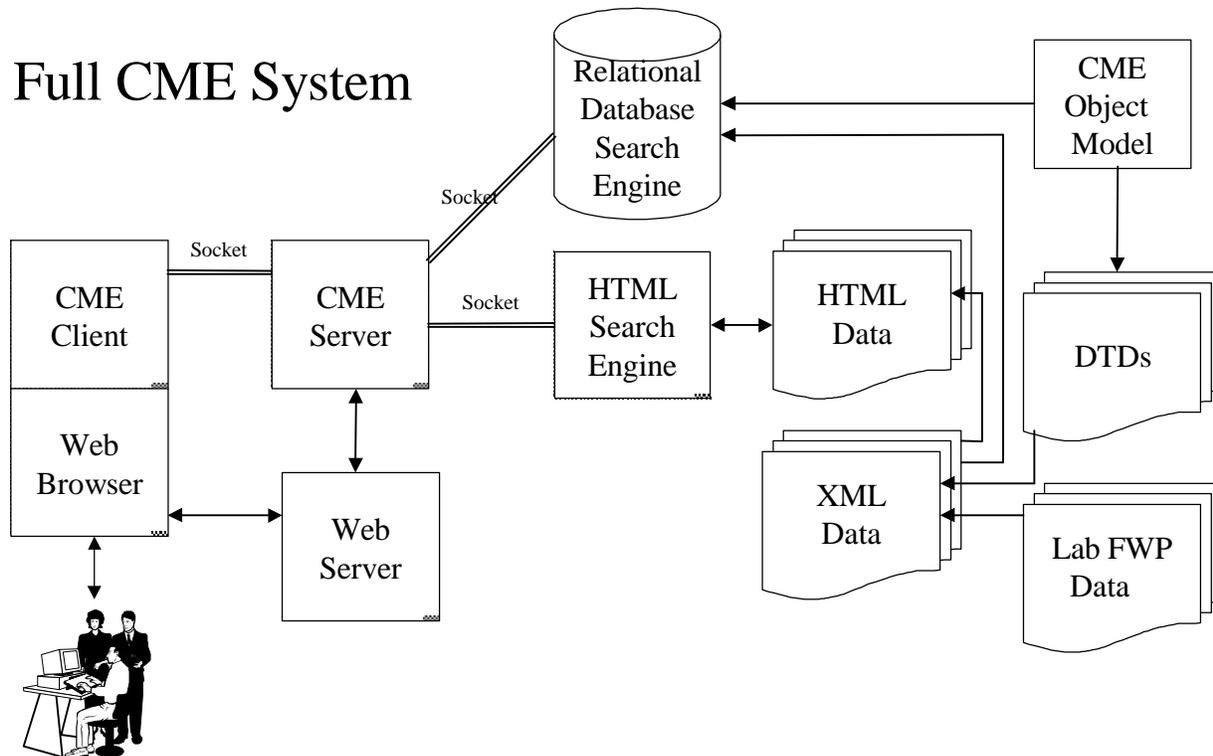
As XML is not currently capable of presenting this type of information, CME currently relies on HTML to provide presentation capability. An XML markup is tagged according to its data content and the structure of the information within the tags, but provides no detail on how to present this information. The eXtended Style sheet Language (XSL) has been defined to provide this presentation capability for XML, however, this technology is emerging and it will most likely be several years before these technologies stabilize and come into wide spread use.

With limited presentation capability in XML, we have developed tools that allow CME to offer a rich information management system using XML, relational database technology, as well as HTML for presentation. In the future, we hope to work towards a pure XML solution. However, from the point of view of a user of CME, the transition to the preferred technology will be transparent.

Functionality

The architecture of this system will become simpler as XML tools mature and stabilize. The key design challenges in this system are to represent XML flat-file information as highly formatted web pages, and to provide keyword and field search capabilities over this information. We chose to translate from XML to HTML presentation format, and to build a relational database from the structure of the XML files to support field queries as shown in Figure 1.

Full CME System



• Figure 1 presents an architectural representation of the CME system and its components.

Figure 1 shows an architectural representation of the CME system and its components. The CME pilot system is a client-server system. The server, *CME Server*, provides the ability to query and generate reports on FWP information. The client, *CME Client*, is responsible for providing the user interface, presenting the information to the user, and relaying user's requests to the server.

The heart of the system, the *CME Object Model*, is not functional, but architectural. The primary objective for the model is to capture all relevant information that appears in a research proposal submitted by a DOE laboratory. This model is then used to define the XML Document Type Definitions (DTDs) for an FWP documents

and to build a relational database for the FWP information. An XML DTD defines the tags that are used to mark up a file. In the FWP DTD file tags such as <TITLE>, <PRINCIPAL INVESTIGATOR>, <PROGRAM MANAGER> are defined, (see Figure 2). From the DTD, an FWP form can be tagged with the appropriate tags. As CME processes thousands of FWPs we have built tools that automatically take FWP information from the laboratories, build the appropriate XML files from this information, and populate the relational database. The CME Server acts as a front end to this data by providing the interface for searching, and directing the Web Server to present the desired web pages to the user's Web Browser.

```

<!ELEMENT cme-proposal (proposal-rec-item)+>

<!ELEMENT proposal-rec-item
  (proposal |
   staff-resources |
   funds-resources |
   equipment-resources |
   budget-attachment |
   equipment-attachment |
   major-cost-attachment |
   time |
   milestone-task |
   doe-program |
   doe-program-manager |
   oo-reviewer |
   pi |
   proposal-manager |
   hq-organization |
   operations-office |
   attachment |
   contractor |
   funds-actual |
   cost-actual |
   report-requirements
  )+>

<!ELEMENT proposal
  (ben |
   work-prop-no |
   work-prop-title |
   rev-no |
   inst-plan-flag |
   work-prop-desc |
   fwp-rpt-rqmt
  )+>

<!ELEMENT ben (#PCDATA)>
<!ELEMENT work-prop-no (#PCDATA)>
<!ELEMENT work-prop-title (#PCDATA)>
<!ELEMENT rev-no (#PCDATA)>
<!ELEMENT inst-plan-flag (#PCDATA)>
<!ELEMENT work-prop-desc (#PCDATA)>
<!ELEMENT fwp-rpt-rqmt (#PCDATA)>

```

...

• Figure 2 A section of the FWP DTD that was defined for the CME project.

The laboratory-specific DTDs allow for effective electronic submission of FWP forms by each participating laboratory. Ideally, each laboratory FWP submission should already be properly tagged using the appropriate DTD tags. However, this burden is quite large for the participating laboratories, therefore the majority of the data translation work is done by the system developers, not the participating labs. We have

developed a tool that converts general laboratory FWP reports into the laboratory-specific XML files. The CME system design allows for distribution of the XML files across the participating laboratory sites. For the current pilot, the XML files reside at the Oak Ridge National Laboratory on a dedicated machine separate from the CME system.

U. S. DEPARTMENT OF ENERGY		
FIELD WORK PROPOSAL		
PROGRAM:		
Budget Element Number: 19980806121046411		
1. WORK PROPOSAL NO. ERKJ009	2. REVISION NO. 0	3. DATE PREPARED 04/15/1998
4. WORK PROPOSAL TITLE: Collaborative Management Environment		5. BUDGET AND REPORTING CODE: EJ0102000
6. WORK PROPOSAL TERM: //OPEN	Patent Status This proposal is being transmitted in advance of patent review evaluation purposes only. No further dissemination or publication shall be made without prior approval of the Assistant General Counsel for Patents, DOE.	Include in Institutional Plan? Y
8. HEADQUARTERS/OPERATIONS OFFICE PROGRAM MANAGER: John Doe (801) ZZZ-ZZZZ	11. HEADQUARTERS ORGANIZATIONS: Energy Research	14. DOE ORGANIZATION CODE: ER
9. OPERATIONS OFFICE WORK: Jim Doe (423) ZZZ-ZZZZ	12. FIELD OFFICE: Oak Ridge Operations	15. FIELD OFFICE CODE: ON
10. PRINCIPAL INVESTIGATOR: Jane Doe (423) ZZZ-ZZZZ	13. CONTRACTOR NAME: Oak Ridge National Laboratory Managed by Lockheed Martin Energy Research Corporation For the U.S. Department of Energy Post Office Box 2008 Tennessee 37831-8085	16. DOE CONTRACTOR CODE: 41
17. WORK PROPOSAL DESCRIPTION: The Collaborative Management Environment (CME) is a joint research project between the Oak Ridge National Laboratory (ORNL) and Lawrence Berkeley National Laboratory (LBNL) that is funded by the U.S. Department of Energy (DOE). CME is designed to support distributed and centralized environments. The goal is to provide access to proposal and project information of importance to DOE and its contractors by establishing ...		

Figure 3 Rendering of an HTML file generated from XML proposal submission

Results

Our preliminary evaluation of the system was focused on the cost efficiency of the approach when adding a new laboratory participant to the system. We chose a very large, multipurpose national laboratory for our evaluation. It took approximately 2 days of a skilled person's time to retrieve and format the data we needed. This is a fairly short amount of time, particularly when compared with other alternatives, such as relational databases or object request brokers. Preliminary results with an additional large defense laboratory show similar results.

The tools that we developed to translate the lab's raw FWP information into XML files must be flexible enough to work with a variety of potential report formats. We accomplished this by building a general-purpose translator from a very general series of reports to a specific XML file. This allows the lab to develop a simple database query to retrieve the desired data. To summarize the effort, it took two days for a lab expert to generate the FWP information, three weeks to

develop translation tools to convert general database reports into XML and HTML files, and about a week to integration the new XML information into the CME system. The CME Pilot functioned as expected, requiring very little lab expense, while maintaining strong functionality.

Discussion

We are using XML as a storage layer definition language enabling distributed information management. This is clearly beyond what XML was originally designed to accomplish. From a developer's standpoint, it can be argued that such an XML-based system has the potential to provide an important component of the functionality of a distributed database for a fraction of the cost. We believe that the key issue addressed by our XML-based solution is meeting the user's "quality of service" needs. By quality of service, we mean providing a solution that meets the users needs at a cost the user can afford. For example, a user of the CME system does not need many of the features that distributed relational database solutions offer. Nor are the providers of the data

willing to pay for the creation and maintenance of such as system. The key question we addressed is not what was the best technology, but rather what is the most cost-effective technology to solve this problem.

We have demonstrated that XML has great potential for managing structured, distributed data, however, there are drawbacks to XML at this time due to the immaturity of the technology. We found that books on XML are often out of step with the tools they reference. The tools are immature, tending to be error prone and limited in scope. Consequently, we processed XML with a combination of currently available XML tools and Java applications. Certainly, as XML tools evolve and mature this will become a mute point, however, for the near term the immaturity of this technology is a noteworthy limitation of XML.

Immaturity aside, we believe that XML has a great potential to become a key future technology. It has the strengths of providing structured data in a distributed manor and is fairly easy to use. A key future issue for XML is how to balance functionality with ease of use. In our case, the simplicity of XML made the technology viable for our application. However, had XML been more complicated then it might suffer the high cost of implementation that many other technologies now face.

Future Trends

An *extended* pilot of the CME system is currently underway. The plans include expanding the system to include three new labs. It is expected that with the addition of each new lab, lessons learned will extend the flexibility of the system, while continuing to minimize impact to the laboratories. Additionally, the CME system is being evaluated for a broader role within DOE.

Our vision for CME is to capture data and its content information in an XML format, then capture the data presentation information in an XSL format, which would allow the Principal Investigator control over that presentation view. Due to the relative newness of XML/XSL approach, and that few browsers support XSL natively we have developed tools that approximate an XSL-based conversion to HTML. Rather than serving an XML file and associated XSL style

sheet to an XSL enabled browser, this converter generates static HTML representations of the XML from XSL directives. Alternatively, the labs may provide the URLs of lab built HTML representations, offering the PIs greater control over an FWP presentation. These static HTML pages could then served to the currently available HTML browser technology, see Figure 3.

The main issue that we face in expanding this pilot to additional labs is that labs will not be able to provide data in a format that we desire. The current FWP system is paper-based. Each laboratory has developed its own system for generating the paper. Some gather full FWP data into a database, and then use database report writers to generate the paper. Others have an informal system with each Principal Investigator generating his FWPs in his or her own preferred manner, i.e., using a favorite word processor, or tagging language. This implies that the only source of electronic FWP information may lie with each PI. Gathering this data from each PI is quite a large task that may have a simple solution. This solution would require the PI to convert his or her FWP to an HTML format when the FWP is submitted for approval. This would allow the FWP information to be displayable from the CME system, also allowing the PI to potentially have a great deal of control over how the proposal looks to potential sponsors. Once an FWP is in HTML, the PI would be free to enhance the proposal with graphics, audio, video, or a variety of other technologies. Our long-term desire would be to have this FWP information represented in XML.

Summary

Through the development of a pilot system to strengthen the research proposal management system at DOE, we have shown that XML provides a simple, low-cost means of distributed information management that is typically performed by a large-scale enterprise information systems. The system requirements called for the proposal information to be distributed, and to incur a minimal cost to the owners of the information. Therefore, the development of a low-cost, distributed information management system is required.

Distributed databases, object request brokers, and various middleware packages deal

very nicely with distributed, heterogeneous data, however, at quite a large cost to the owners of the data. Our approach is to use XML as an information repository that can be used to search for and present information. The produced XML files can be seen as a “database” of proposal information. We developed converters to generate raw database reports into XML files, various means of querying this information, including a tag-based search engine, and various ways of presenting this information, for example XSL style sheets.

Applying this approach to a large, multipurpose national laboratory with hundreds of research proposals demonstrates the very low cost of integrating their information into our system, while meeting the initial system requirements. Use of XML as a low-cost means of representing distributed and structured data is certainly feasible, and depending on constraints, may provide the only viable solution. XML is an emerging technology, and suffers from immature tools, and supporting standards. These shortcomings will no doubt be addressed over time. XML appears to have a very bright future.

Acknowledgements

The Oak Ridge National Laboratory is managed by Lockheed Martin Energy Research Corp. for the U. S. Department of Energy under contract number DE-AC05-96OR22424. This work was supported

by the Department of Energy’s Division of Mathematical, Information, and Computational Sciences (MICS) Office. We would like to thank Dan Hitchcock for his continuing support of the CME project, and Kimberly Barnes, the original Principal Investigator of the CME Project.

References

- Bray, T., Paoli, and Sperberg-McQueen, C. M. ‘Extensible markup language’ (XML) 1.0. A W3C standards committee recommendation (February 1998). Available at <http://www.w3.org/TR/REC-xml>
- Bosak, J. ‘XML, Java, And The Future Of The Web.’ *World Wide Web Journal Volume II*, Issue 4 (Fall 1997)
- Lincke D. M. and Schmid, B. ‘Mediating electronic product catalogs.’ *Communications of the ACM*. 41, 7(July 1998), 86-88.
- Singh, N. ‘Unifying heterogeneous information models.’ *Communications of the ACM*. 41, 5(May 1998), 37-44.
- Potok, T. E., Ivezic, N, Singletary, B. A., “XML For Web Based Collaboration.” *Proceedings of XML98 Conference*, Chicago, IL, 1998.