

Electronic Publications Pilot Project

Final Report

June 2001



Ohio Historical Society



Ohio Supercomputer Center



State Library of Ohio

Electronic Publications Pilot Project

Contents

1.	Executive Summary	Page 4
2.	Introduction	Page 5
3.	Status of Project Deliverables	Page 6-7
4.	Project Procedures	Page 8-19
5.	Conclusion	Page 19-20
	Appendix A – Contract between SLO & OSC	Page 21-23
	Appendix B – “Completed software & system criteria table	Page 24-26
	Appendix C – Target Web Publications	Page 27
	Appendix D – Searching & Harvesting Metadata Elements	Page 28

Electronic Publications Pilot Project

Team Members

Jim Buchman

Head of Public Services

State Library of Ohio

274 E. 1st Avenue

Columbus, OH 43201

614-644-6954

jbuchman@sloma.state.oh.us

Gretchen Persohn

Head of Research Services

State Library of Ohio

274 E. 1st Avenue

Columbus, OH 43201

614-644-5952

gpersohn@sloma.state.oh.us

Kathy Hughes

Technical Services

State Library of Ohio

274 E. 1st Avenue

Columbus, OH 43201

614-728-8320

khughes@sloma.state.oh.us

Judy Walker

Assistant State Archivist

Ohio Historical Society

1982 Velma Avenue

Columbus, OH 43211

614-297-2585

jwalker@ohiohistory.org

Nicole Merriman

Government Information Services

State Library of Ohio

274 E. 1st Avenue

Columbus, OH 43201

614-995-4117

nmerriman@sloma.state.oh.us

Kevin Wohlever

Senior HPC Systems Specialist

Ohio Supercomputer Center

1224 Kinnear Road

Columbus, OH 43212-1163

614-292-4167

kevin@osc.edu

1. Executive Summary

With this report, the Electronic Publications Pilot Project (EP3) team concludes its work as stipulated in the personal service contract between the State Library of Ohio (SLO) and the Ohio Supercomputer Center (OSC), which is included with this report as Appendix A. In addition to staff from OSC and SLO, staff from the Ohio Historical Society (OHS) played an integral role in this project.

The purpose of this project was to develop a process to identify, capture (or harvest) and permanently store in a searchable database, web publications created by state of Ohio agencies. By identifying issues, developing criteria, reviewing off-the-shelf software applications, and preparing training materials, EP3 would provide a jumpstart to the Joint Electronic Records Repository Initiative (JERRI).

The EP3 team spent considerable time investigating this issue. Some of our general discoveries are noted below.

- No off-the-shelf software application currently exists.
- Any software application will require extensive monetary resources to develop and implement a workable solution.
- The magnitude of the problem, and its long-term implications of lost information of enduring historical significance, is enormous.
- Current state of Ohio web sites lack necessary descriptive data (metadata) needed to accurately identify significant publications.

Based upon our study, the EP3 team makes the following recommendations:

- That no additional funding be allocated to purchase and implement any of the products evaluated at this time.
- That a partnership be formed with OCLC, at this early stage, to advise on the development of the OCLC Digital Vault Project.
- That the JERRI partners (Department of Administrative Services, Ohio Historical Society, Ohio Supercomputer Center, and the State Library of Ohio) continue working together to implement a long-term solution.

2. Introduction

2.1. Project Overview

The Electronic Publications Pilot Project was designed to be an interagency cooperative venture with OSC, OHS, and SLO.

EP3 began to address the issues of storage and access to State of Ohio government information and publications produced in an electronic format. The original project plan was to harvest information and publications from state agency web sites, using customized software, and store the data at OSC. SLO and OHS staff were to review harvested data to determine which publications met established criteria for long-term storage. SLO staff would then create catalog records in its on-line catalog for those publications meeting established criteria. The SLO catalog record would provide a direct link to the information, located at OSC, and would also be available through the OhioLINK network.

As a pilot project, EP3 was to build the foundation that would become the basis for JERRI (Joint Electronic Records Repository Initiative). JERRI will create elaborate search engine, index and permanent storage of information harvested through EP3. Funding for JERRI was requested by DAS (Department of Administrative Services) and OHS in their state FY 02-03 budget requests; however, these funding requests were not included in the final FY 02-03 biannual budget.

LSTA funds were to be used to purchase the necessary hardware and software needed for EP3, and to develop training materials, including an on-line tutorial, aimed at agency personnel responsible for adding information to their agency's web site.

A project team was established to oversee development of EP3. This team was chaired by SLO, and also included three SLO staff as members. Also, the project team included one staff member from OHS, two from OSC, one from DAS Office of Policy & Planning, and one from OhioLINK.

The project began upon receipt of a contract from SLO to OSC, approximately the beginning of December, and will end on June 30, 2001. This contract included a list of Project Deliverables, which is used as a basis for this report.

3. Status of Project Deliverables

3.1. *Consult on issues of access to and preservation of web based State of Ohio government records and publications.*

- OSC has provided, and will continue to provide, technical and policy support on issues of access to and preservation of web based State of Ohio government records and publications.

3.2. *Define software criteria, capabilities and requirements, including:*

3.2.1. *Automated identification of web based publications created by Ohio State government.*

3.2.2. *Review of web based publications to determine which publications meet criteria for long-term preservation, storage and access.*

3.2.3. *Automated retrieval and storage of selected Ohio State government web based publications.*

3.2.4. *Automated indexing, metadata harvesting and classification of web resources using a web crawler for current state resources.*

- OSC and the EP3 team developed software and system criteria, capabilities and requirements for a system to do automated identification, review, indexing, storage, and retrieval of web based publications created by Ohio State Government. A list of the criteria and sample evaluation sheet used to compare vendors and products is included in Appendix B at the end of this report.

3.3. *Develop training materials, including an on-line tutorial, aimed at agency personnel responsible for their agency's web site.*

- This deliverable could not be completed. Since software acquisition was delayed, and currently unavailable, training materials could not be developed. This deliverable was removed at the recommendation of the SLO EP3 project lead since no off-the-shelf "product" existed from which an on-line tutorial could be developed.

3.4. *Evaluate software and recommend product(s) to purchase for the EP3 project.*

- With this report, OSC and the EP3 team have completed this deliverable.

EP3 Final Report

- 3.5. *Purchase, install and configure hardware and software for EP3 project.*
- With this report, OSC and the EP3 team describe this completed deliverable and its product(s) status.
- 3.6. *Assess resources required and recommend processes for addressing the identification, selection and preservations of State of Ohio web publications created prior to the implementation of these tools.*
- With this report, OSC and the EP3 team have completed this deliverable.
- 3.7. *Provide input into JACARR rules to address permanent storage of electronic publications as part of the proposed changes to ORC 149.11.*
- OSC has provided, and continues to provide input into JCARR rules to address permanent storage of electronic publications as part of the proposed changes to ORC 149.11.
 - State Library staff have drafted changes to the language in ORC 149.11 to include electronic publications. Staff have not yet begun the legislative process to effect an actual change to the code.
- 3.8. *Create catalog records of selected web based publications and distribute via SLO online catalog and OhioLINK network.*
- This deliverable could not be completed. Blue Angel Technology was able to provide Dublin Core as a metadata standard for use with the data entry function of MetaStar Repository. This should have allowed the creation of catalog records that could be exported into the SLO online catalog and the OhioLINK network. Due to time considerations, and problems with training, we were unable to experiment with this feature.
 - A second issue identified was that records harvested by the software did not include the necessary data elements to easily create catalog records.

4. Project Procedures

4.1. *Team development and makeup*

The EP3 project team was assembled using the process described in the original contract (Appendix A).

Creation of a Electronic Publications Project Team (EPPT)

- a) SLO to chair and have three additional members
 - project lead - Jim Buchman
 - cataloger - Kathy Hughes
 - librarians - Nicole Merriman, Gretchen Persohn,
- b) Two members from OHS - Charles Arp, Judy Walker
- c) Two members from OSC
 - a) Technical - Kevin Wohlever
 - b) Policy - Sol Bermann
- d) Member from DAS - Dave Larson
- e) Member from OhioLINK - Charly Bauer

A subset of the main project team lead the process of setting up the evaluation criteria, installing the hardware and software, and working with the vendors to understand and evaluate the product offerings.

This subset team consisted of the following working group

- a) From SLO
 - Jim Buchman
 - Kathy Hughes
 - Gretchen Persohn
 - Nicole Merriman
- b) From OHS
 - Judy Walker
- c) From OSC
 - c) Kevin Wohlever

This "criteria" team, as it was initially called, met frequently and used e-mail to exchange ideas, meet with vendors and discuss issues. This group led the development of the reports and recommendation herein contained. The full project team, which included members from OhioLINK and DAS, never met since a fully viable test environment was not available.

4.2. EP3-Criteria Team Results

4.2.1. Development of evaluation criteria

Soon after the EP3 team was assembled, the development of the software evaluation score sheet was begun. Key functions were identified. Once the functions were identified, they were sorted into broad categories of "required" or "nice".

A "required" feature is one that the team felt any solution must contain in order to be considered for further evaluation.

A "nice" feature is a feature that would be nice to have, but would not eliminate a solution if the feature were not available.

Below are the functions and categories that the team developed.

Function	Requirement Category
Special Notes	
Centralized Server - The search must run and be controlled from a single system.	Required
Responsiveness of software company - Can we get the proper support?	Required
Customization opportunities of software - If it does not do it all now, is the company willing to modify, or are there customizable options?	Required
Wildcard targeting of web sites - The need to select web sites to harvest without having to identify them all at start time.	Required
Identification of object (file) type - What type of object is being harvested.	Required
Identify, harvest and notify items that have been changed - Based on update date or some type of id check.	Required
Selection flexibility - Must be able to harvest many types of object or files.	Required
Selection review opportunity – Can we review selections before automatic inclusion into the archive?	Required
Reporting flexibility - Must be able to report on harvest information, amount of data, harvest location, etc	Required

Function	Requirement Category
Cost – Lower costs are nicer to work with.	Nice
"Crosswalk" to MARC format – It would be nice to be able to automatically catalog the harvested object in the proper library catalog format.	Nice
Location of software company – Companies that are closer to Columbus or have offices near Columbus may be easier to work with	Nice
Ability to customize spider identification to webmaster to avoid problems with targeted servers.	Nice
Customization of software provider - Will the project team be expected to do the customization? It is nicer if the vendor does the coding. It is not quite as nice if a third party must do the modifications.	Nice

EP3 Final Report

4.2.2. *Vendor /product evaluation*

The EP3 team developed a list of vendors to contact about possible solutions. The vendors and products to be evaluated were suggested by colleagues or were identified through professional organizations and personal contacts.

Six products and vendors were identified for evaluation by the EP3 team. The products that were evaluated were:

- Web Tracker from Intelliseek
- CORC from OCLC
- MetaStar Suite from Blue Angel Technology
- Site Server from Microsoft
- Compass and Grapevine from Netscape
- eoExchange from eoExchange

Gathering detailed information about the products, from knowledgeable sources, proved to be more difficult than identifying the products. One vendor, eoExchange based in San Francisco, CA, never responded to repeated requests for additional information. Attempts for contact included e-mailing general sales from their web site, e-mail and calls to their VP of Marketing. Needless to say, their product was dropped from further evaluation.

Most other vendors were almost as difficult. OCLC, Microsoft and Netscape only worked through consultants or resellers. It took much longer to arrange the resources we required by following these requirements.

Two vendors, Blue Angel Technology and Intelliseek provided knowledgeable technical presentations at OSC.

Initial planning had the initial vendor evaluations being completed by early December. Due to the problems of lining up the proper resources from the vendors, the initial evaluations were not completed until mid-January.

4.2.3. *Changes to deliverables based on costs*

Once the information from the initial vendor evaluations was obtained and reviewed, it became apparent that the initial plan of obtaining 2 or 3 software packages and running them in a side by side comparison would be difficult. For a short period of time, an effort was made to set up a side by side comparison between two products. However, before both products were in place it became clear that one of the products was not going to meet the needs of the project.

Consequently the team requested and was granted a change in the original project plan so that only one product could be evaluated.

The most critical problem was that very few of the vendors seemed to have a viable product that we could test at this time. Most of the products required extensive modifications, normally by a third party consulting firm, in order to meet the project requirements.

Although the project requirements were that unusual most vendors had yet to implement the strategy EP3 is taking.

4.2.4. *Software recommendation*

After reviewing the capabilities of the products and how well the products met the requirements of the EP3 team, it was determined that two products met most of the criteria to be the long-term solution for JERRI.

One product was the MetaStar software suite from Blue Angel Technology. This company is located in Blue Bell, Pennsylvania. The MetaStar product suite has been selected for use within a number of state library systems. The product suite has a number of the required JERRI features available today. The evaluation would be completed with the assistance of the vendor.

The second product suite selected for evaluation was the combination of the GrapeVine Software and the iPlanet products. At the time GrapeVine was in the process of being acquired by iPlanet. I Planet is an E-commerce solution brought forth by the alliance of Sun Microsystems and Netscape. The evaluation would

EP3 Final Report

be completed with the assistance of a consulting group, MRK Technologies, based in Cleveland, Ohio.

4.2.5. *Software evaluation – Results*

Shortly after beginning the process to do the parallel software evaluations, MRK Technologies, the consulting group assisting us with the iPlanet software, informed the EP3 team that although they felt the GrapeVine, iPlanet software could provide the solution the team was looking for, it would take significant programming and financial resources to modify the software to provide a starting point for the product evaluations. After some discussion with the project team, it was decided that this solution evaluation should not be pursued further.

This allowed the EP3 team to concentrate their efforts on the MetaStar product suite evaluation. Although this decision limited the number of products to evaluate, it strengthened the overall project because preparing for and completing the evaluation required a large amount of resources and staff time that would not have otherwise been available.

The software, process and procedures that the EP3 team would be evaluating would involve the following features and / or requirements:

- Identification of materials on web sites and harvesting of the materials for inclusion in the EP3 archive.
- Tagging harvested materials with metadata to ease future retrieval of the information.
- Storing the harvested information and the associated metadata.
- Secure, authorized access to the stored material.
- Presentation of the stored material to users via a web browser.

Preparation for evaluation of the Blue Angel Technology (BAT) MetaStar software required setting up three systems, with a number of software installations. The BAT MetaStar software could be configured in a Microsoft NT only configuration, or a Solaris / Unix configuration. The EP3 team chose to install the

EP3 Final Report

Solaris / Unix suite as execution in a Unix environment was a requirement determined early in the project.

Preparation of the BAT environment was slowed by the lack of training for the system administrator, in the software.

Unfortunately, the environment needed to be configured before training could take place.

There were three systems, dedicated for use of the EP3 team, to evaluate with the BAT MetaStar product suite. The configuration and function of the systems will be described in the next few paragraphs.

The bulk of the BAT software was installed on a Sun Ultra 3000 running Solaris 5.8. This system had 6 UltraSparc II CPUs running at 248 MHz each and 768 MBs of memory. The MetaStar software installed on this system included:

- MetaStar Server – Provided access to the stored data.
- MetaStar Data Entry – Provided bulk data insertion to the database.
- MetaStar Gateway – Provided web browser access to the information provided by MetaStar Server.
- MetaStar Harvester – Traversed the network and identified and harvested electronic documents (brought documents back to the server).

In order for the MetaStar software to function properly some additional software was required on the system. The first additional product was the Apache web server. This served the web pages generated by the MetaStar software to users. Another product used was called JRUN, which served java applets to run on web browser.

The rest of the BAT MetaStar software, Repository, was installed on an NT workstation. The workstation used for this was a SGI Visual Workstation 320, running Microsoft NT 4.0. Once again, additional software was required in order for the BAT software to have communication capabilities and full functionality. On this system the additional software was Oracle Client Tools 8.0.6 and software called "Data Direct" from Merant Software. These

EP3 Final Report

additional packages allowed the NT system to exchange data with the Oracle backend storage repository.

The backend repository required Oracle to be running on a system. For this function it was decided to run the Oracle software on a SGI Origin 200 running IRIX 6.5.10. Oracle 8.0.6 was installed on this system. The Oracle database could reside on any system that supported Oracle. Our system selection allowed for a system to be dedicated to the evaluation process.

Once the systems were configured, training of the EP3 team and eventual evaluation of the BAT software could begin.

The original plan had the EP3 team being trained in Columbus by a trainer from BAT. The training was to take 4 days to complete. Unfortunately, this plan had to be changed at the last minute due to some medical complications involving the trainer.

Instead, the team was given training by the instructor through a series of conference calls. This training allowed the EP3 team to search State of Ohio web sites, identify documents, harvest documents, apply metadata to documents, archive the documents, and later present the data.

All in all, the BAT MetaStar software suite functioned as advertised, but not necessarily as expected by the EP3 team.

One of the unexpected issues that came to light during the evaluation was the importance of the BAT Repository, residing on a single NT workstation. All harvested documents had to be transferred to the NT system where the BAT repository software would move them into the Oracle database. The NT workstation was also the only place to apply the metadata.

The requirement to pass all the information and data through the NT workstation proved to be a major logistical problem. The movement and eventual archiving of documents required administrator level access to the data. Unfortunately, providing the metadata also required this level of access. This is a security access issue.

The lack of network access capabilities under Microsoft NT required that any user, such as a cataloger applying metadata, be

EP3 Final Report

at the NT console. This caused problems when the system administrator and cataloger were located in different parts of the city.

There were also a couple of issues with harvesting the electronic documents. The BAT software, normally, identified the web location of the document and returned that information to the harvest server. The EP3 team wanted the documents returned and placed into the archive.

The EP3 team also wanted the document URL to point to the location in the EP3 archive. Unfortunately, this capability was not available, although BAT noted that it would not be hard to modify the software to support this function.

The team also was able to get some information on the resources and skill sets needed to support the complete installation of the BAT software suite.

To support the BAT MetaStar software suite, a system administrator with Sun Solaris UNIX and Microsoft NT skills is needed. A strong understanding of system and network security issues is also required.

This administrator will need to be able to dedicate 100% of their time to the initial installation of the systems and software. After which, the systems would expect to take 25 to 50% of the administrators time, depending on the size of the archive and the number of users and the security procedures that must be supported.

In addition, the support of a knowledgeable Oracle database administrator would be very beneficial. A person with Oracle administration skills could expect to spend 20% of their time supporting the BAT environment, after the initial installation.

4.2.5.1. Security / Network Resources

The harvesting aspect of the system requires that a good network connection be available in order to crawl the web and send back the harvested information.

This data transfer back to the host will need to be further investigated for its impact on site security. A minimum of 50

GBs of disk holding space is necessary to hold the harvested documents before placing them in the repository.

The overall security of the system was not evaluated during this project. Project security is something that needs to be evaluated in a separate project, as it impacts many levels of the host site security, and the access to the data.

4.2.5.2. *Capital Funding*

Capital funding requirements for this project are broken down into the following parts: Computer Hardware, Computer Software and Networking.

One of the appealing features of the BAT MetaStar solution, is the flexibility of the location of the software installation. All of the Unix software can be located on one system, or each piece installed on a separate system. The NT based repository software must be located on the NT server.

The model that we used to start with seems to be a good starting point for configuring the rest of the system. Having a single UNIX system to support the harvesting, metadata work, and presenting data back to users works well. Especially if the procedures call for harvesting data in the evenings and weekends, and the metadata work and the bulk of presenting data occurring during a normal work day.

Once the workload becomes high enough, moving harvesting and then data presentation to their own servers will reduce the workload on the initial system, and increase productivity of the systems and users.

Finally, a separate, easily expandable, highly dependable system needs to be in place to support the back end database archive. In our initial planning, we hoped to use the NCR teradata system, but the Blue Angel Technology software only supported Oracle or Microsoft Access.

In the following table, approximate costs for purchasing the software and hardware are listed. Estimated annual maintenance fees are also shown. No attempt was made to

EP3 Final Report

estimate the personnel costs, system security or the network connectivity costs.

Item	Purchase Price	Annual Maintenance Costs
Software		
Blue Angel Technology, MetaStar Software	\$100,000	\$15,000
Oracle Database Software	\$100,000	\$20,000
Merant ODBC Driver	\$200	-0-
Hardware		
NT Workstation	\$6,000	\$100
Harvest / Metadata System	\$30,000	\$3,000
Database Server	\$150,000	\$10,000
Total for H/W and S/W	\$386,200	\$48,100

4.2.5.3. *Web publication information*

Web publications of State of Ohio agencies, though a subset of all state records, were the focus of EP3. See Appendix C for a listing of the types of web publications targeted. There are approximately 145 web sites for State of Ohio agencies (<http://www.state.oh.us/agencies.htm>). Due to reasons cited in the conclusion, the total size, frequency of updates, and storage requirements could not be ascertained.

4.2.5.4. *Metadata requirements / suggestions.*

One of the first things that the EP3 team did, was to define which metadata elements were essential (see Appendix D). Once these elements were defined, they could be used to help catalog the electronic publications, as a basis for training web masters, and, ultimately, to help end users find what they are looking for.

Once the team started product testing, it was found that the harvested publications had very little, if any, metadata attached. This meant that someone would have to add all of the missing information to each publication before it could be indexed and archived.

The presence of metadata is also critical in cataloging. Currently, OCLC provides the CORC service, which allows for the cataloging of

EP3 Final Report

websites and electronic publications through the automatic transfer of metadata onto bibliographic templates. Again, any metadata that is not attached to the publication or website must be supplied manually by the cataloger.

Clearly, webmaster training and support in the addition of metadata to documents is critical before automatic indexing and cataloging of electronic publications is possible.

4.2.5.5. *Web Master Awareness/Training Issues*

It was clear from the harvested records reviewed by the group that state agency web pages lack descriptive metadata. This information is crucial for cataloging, searching, storing and retrieving publications. In an ideal world, this descriptive information would be added when the electronic publication is created, either by the web master or by the publication's author. Otherwise, excessive staff time would be required to manually review harvested data and create metadata.

This is an awareness issue that must be raised with state agency web masters. One venue for doing this is through the OERC (Ohio Electronic Records Committee).

5. Conclusion

Although the project was started with great enthusiasm and high hopes, the reality of the current situation for harvesting and archiving of electronic documents eventually required the EP3 team to reassess their goals and reduce the expectations.

There is not a hardware and software solution currently available that meets the current, let alone future, requirements for electronic document archiving in the State of Ohio. Current solutions are also expensive, based on current budgets, to obtain, implement and support.

Although the MetaStar software suite from Blue Angel Technology showed much promise, it also had its drawbacks. That solution will work well for centrally located entities that will harvest, apply metadata and archive electronic documents. In its current state, the BAT solution would not work well in the distributed model preferred by the consortium of the Ohio Historical Society, the State Library of Ohio and the Ohio Supercomputer Center, and renders the BAT solution unworkable.

EP3 Final Report

But the news is not all bad. Governments and solution providers are beginning to recognize the problem and apply more resources to possible solutions. Some of the so called "heavy hitters" are now stepping up to provide solutions that not only work, but interact with other archive and catalog systems, allowing the state to preserve and enhance investments in on-line catalog systems.

One of the potential solutions is being generated in central Ohio at the Online Computer Library Center (OCLC). OCLC is in the process of partnering with forward looking groups in an effort to develop an electronic document archive solution.

- It is the recommendation of the EP3 team, for the many reasons expounded upon in the preceding pages, that no funding be allocated to the purchase and implementation of any of the products evaluated.
- It is the further recommendation of the EP3 team, that a partnership be formed with OCLC, at this early stage, to advise on the development of a process to identify, harvest and archive electronic documents on the web. Information obtained during the course of the Electronic Publications Pilot Project can be used to develop a solution viable for the state of Ohio and other government entities.

EP3 Final Report

Appendix A - Contract between SLO and OSC for EP3 project.

THE STATE LIBRARY BOARD

AGREEMENT FOR SERVICES

This AGREEMENT is entered into this ___ day of December 2000 by and between the Ohio State Library Board, hereinafter known as the State Library, located at 65 South Front Street, Columbus, Ohio 43215-4163, and The Ohio State University, on behalf of the Ohio Supercomputer Center (OSC) hereafter referred to as Consultant.

NOW THEREFORE, in consideration of the promises hereinafter set forth, both parties mutually agree as follows:

ARTICLE I - PURPOSE

(A) To coordinate the Electronic Publications Pilot Project (EPPP) in conjunction with the State Library of Ohio and the Ohio Historical Society.

ARTICLE II - STATEMENT OF SERVICES

- (A) The Consultant, in cooperation with the EPPP Team, will perform the following services:
- 1) Consult on issues of access to and preservation of web based State of Ohio government records and publications.
 - 2) Define software criteria, capabilities and requirements including:
 - a) Automated identification of web based publications by State of Ohio government.
 - b) Review of web based publications to determine which publications meet criteria for long-term preservation, storage and access.
 - c) Automated retrieval and storage of selected State of Ohio government web based publications.
 - d) Automated indexing, metadata harvesting and classification of web resources using a web crawler for current state resources.
 - 3) Develop training materials, including an on-line tutorial, aimed at agency personnel responsible for their agency's web site.
 - 4) Evaluate software and recommend product(s) to purchase for the EPPP Project.
 - 5) Purchase, install and configure hardware and software for the EPPP Project.
 - 6) Assess resources required and recommend processes for addressing the identification, selection and preservation of State of Ohio

EP3 Final Report

web publications created prior to the implementation of these tools.

- 7) Provide input into JCARR rules to address permanent storage of electronic publications as part of the proposed changes to ORC 149.11.

ARTICLE III - TIME OF PERFORMANCE

- (A) This Consultant's services shall commence upon receipt by the Consultant of an approved and signed copy of this contract and approval of the contract by the Ohio Office of Budget and Management and end on June 30, 2001.
- (B) This contract shall be binding on both parties upon receipt by the Consultant of an approved and signed copy of this contract. Provided, however, that it is expressly understood and agreed to by the parties that none of the rights, duties and obligations herein shall be binding to either party, until all statutory provisions of the Ohio Revised Code, including but not limited to Section 126.07, have been complied with and until such time as all necessary funds are available or encumbered.

ARTICLE IV - COMPENSATION

- (A) All compensation will be made to OSC.
- (B) The Consultant will be paid as follows:
 - 1) Development & submission of selection criteria document - \$30,000.
 - 2) Development of training content outline & training prototype - \$8,000.
 - 3) Final report of completed projected - \$20,000.
- (C) Total compensation for services will not exceed \$58,000.00.
- (E) It is mutually understood and agreed that the Consultant is an independent contractor and will receive no fringe benefits and that there will be no money withheld from the Consultant's compensation for income taxes or retirement or for any purpose. The Consultant will carry insurance which will absolve the State Library from any responsibility in case of accident.

ARTICLE V - ACCEPTANCE

- (A) The Chair of the EPPP Team will be the sole judge in determining whether services provided by the OSC are acceptable.

ARTICLE VI - COMPLIANCE WITH LAW

- (A) The Consultant agrees to comply with all applicable federal, state, and local laws in the conduct of the work under this agreement.

EP3 Final Report

ARTICLE VII - EQUAL EMPLOYMENT OPPORTUNITY

- (A) In carrying out this contract, the contractor will not discriminate against any employee or applicant for employment because of race, religion, color, sex, sexual orientation, national origin, handicap or age.

ARTICLE VIII - CONFLICT OF INTEREST

- (A) The Consultant shall not voluntarily acquire any personal interest which is in conflict with the fulfillment of his responsibilities with respect to the carrying out of the work under the contract.

Any person who voluntarily or involuntarily acquires a conflicting personal interest after the effective date of the contract shall immediately disclose his interest in writing. He will not participate in any action affecting the work under this agreement unless the State Library determines that his participation would not be contrary to the public interest.

ARTICLE IX - TERMINATION

- (A) This agreement will terminate on June 30, 2001.

Either party may voluntarily terminate this contract upon notice in writing to the other party.

ARTICLE X - CONSTRUCTION

- (A) This agreement shall be constructed, interpreted, and the rights of the parties determined in accordance with the laws of the State of Ohio.

IN WITNESS THEREOF, the parties have executed this contract as of the day and year first written above.

Signed	Date	Signed	Date
Charlie Bender Executive Director Ohio Supercomputer Center 1224 Kinnear Road Columbus, OH 43212-1163		Michael S. Lucas State Librarian The State Library of Ohio 65 South Front Street Columbus, OH 43215	

EP3 Final Report

Appendix B - "Completed" Software and system criteria for automated identification, review, indexing, storage and retrieval of web based publications created by Ohio State Government.

Function	Requirement Category	OCLC / CORC	Intelliseek / Web Tracker	Blue Angel / MetaStar	Microsoft / Site Server	Netscape / Compass & Grapevine	eoExchange
Special Notes		Talked to Carol Bradsher of OhioNet. 11/21/00	Scheduled for demo 11/30/00	Information sent 11/27/00 from Chris Hynes	Contact with Doug Browne of Dedicated Technology	Contact with Scott Fillmore of iPlanet.	No response to inquiry
Centralized Server - The search must run and be controlled from a single system.	Required		Available	Available	Available	Available	
Responsiveness of software company – Can we get the proper support?	Required		Not Really	Yes	Not Really	Unsure.	
Customization opportunities of software - If it does not do it all now, is the company willing to modify, or are there customizable options?	Required		Can do.	Can do.	Not really. It is script supported and detailed work.	Yes (so they say).	
Wildcard targeting of web sites - The need to select web sites to harvest without having to identify them all at start time.	Required		Yes, available	Available	Available	Available	
Identification of object (file) type - What type of object is being harvested.	Required		Not sure how it works, if it is available.	Possibly.	Not available.	Available	
Identify, harvest and notify items that have been changed - Based on update date or some type of id check.	Required		Yes.	Possibly	Yes	Yes.	

EP3 Final Report

Function	Requirement Category	OCLC / CORC	Intelliseek / Web Tracker	Blue Angel / MetaStar	Microsoft / Site Server	Netscape / Compass & Grapevine	eoExchange
Selection flexibility - Must be able to harvest many types of object or files.	Required		Unsure if it is available.	Yes	Yes	Yes	
Selection review opportunity – Can we review selections before automatic inclusion into the archive?	Required		Yes	Not clear, it looks possible.	Yes, must script.	Yes	
Reporting flexibility - Must be able to report on harvest information, amount of data, harvest location, etc	Required		Yes, some good capabilities.	Available.	Available, but not sure how to use.	Yes, real Time capable.	
Cost - Lower costs are nicer to work with.	Nice						
"Crosswalk" to MARC format - It would be nice to be able to automatically catalog the harvested object in the proper library catalog format.	Nice		Not available.	Possible to implement with XML scripting	Develop local scripts	No information available.	
Location of software company - Companies that are closer to Columbus or have offices near Columbus may be easier to work with	Nice		Cincinnati, Ohio	Blue Bell, PA	Seattle, WA. Consulting support out of Columbus, OH.	Alexandria, VA. Consulting support out of Cleveland, OH	

EP3 Final Report

Function	Requirement Category	OCLC / CORC	Intelliseek / Web Tracker	Blue Angel / MetaStar	Microsoft / Site Server	Netscape / Compass & Grapevine	eoExchange
Ability to customize spider identification to webmaster to avoid problems with targeted servers.	Nice		Yes.	Yes, ID by web address of harvester.	Yes, not a problem	Not a problem.	
Customization of software provider - Will the project team be expected to do the customization? It is nicer if the vendor does the coding. It is not quite as nice if a third party must do the modifications.	Nice		? Unknown.	Done by Vendor.	Customer or consultant.	Customer or consultant.	
Exclusion of Web Sites	Required		No.	No.	No.	No.	
Support of "Framed" web pages.	Required		No.	Yes.	Not sure	Not Sure.	

Appendix C – Target Web Publications

Notes: This listing follows the model of the State Library's current paper requirements for acquiring State of Ohio publications created for the general public. SLO has traditionally received many of the following types of materials. The list is in no particular order and is not exclusive.

- Annual Reports
- Databases (a.k.a. directories)
- Pamphlets/Brochures/Fact Sheets (a.k.a. General Publications)
- Laws & Rules
- Handbooks/Manuals
- Technical/Research/Statistical Reports
- Bulletins/Circulars
- Forms
- Newsletters
- Press Releases
- Maps/Charts
- Multi-media files(sound, video, interactive)

Appendix D - Searching & Harvesting Metadata Elements

Required Fields

- Title (of publication)
- Author (s)
- Originator + Subfields (Publisher)
- Agency Program
- Government type (State, Local)
- Dates
- Creation Date
- Revision Date or Date of Last Modification
- Medium/Object/File/Format Type
- Language
- Public Access Tag (To ID publications intended for general public use)
- Identifier/URL
- Description (Free text field)
- Keywords/Subject Terms (SLO staff can work with individual state agencies to help them develop a list that is particular to their agency)
- Relation (Used for publications that are too large for one file. Relates back to Table of Contents. May require multiple URLs, e.g. Ohio Revised Code or JACAR rules)
- Contact person (s)