

The 20th Annual National Environmental Monitoring Conference



Monday, July 19th — Friday, July 23rd

Proceedings

**Conference: Monday, July 19th – Wednesday, July 21st
Short Courses: Thursday, July 22nd – Friday, July 23rd**

**Wyndham Washington Hotel
1400 M Street, NW
Washington, DC**



**QUALITY TRAINING FOR THE
LABORATORY COMMUNITY**

NEMC is managed by the Independent Laboratory Institute (ILI), in association with the American Council of Independent Laboratories (ACIL) and Instant References Sources, Inc. under a cooperative agreement with the U.S. Environmental Protection Agency. ILI is the educational foundation affiliated with the American Council of Independent Laboratories.

USING SEDD FILES WITH THE WEB-BASED ENVIRONMENTAL INFORMATION MANAGEMENT (EIM) SYSTEM



Todd M. Pierce

Locus Technologies, 1333 N. California Blvd., Suite 350, Walnut Creek CA 94596

piercet@locustec.com

Overview

The Environmental Information Management (EIM) system is a web-based application developed by Locus Technologies that lets users manage, query and report their analytical and geotechnical data. EIM is used to manage analytical, field, survey, geologic and other environmental data for almost 4,000 sites.



Figure 1. The EIM System from Locus

To take advantage of web-based data management systems such as EIM, analytical laboratories need a way to deliver data that can be understood by any web based system. One data delivery format that meets this requirement is the Staged Electronic Data Deliverable (SEDD) format. By generating data in SEDD format, an analytical laboratory ensures that its data can be used by any web service or XML-based web application. The analytical data 'life cycle' can thus be made quicker and more efficient from sample collection through lab analysis to final report preparation.

The SEDD Format and XML

The Staged Electronic Data Deliverable (SEDD) is an inter-agency effort spearheaded by the U.S. Environmental Protection Agency (EPA) and the U.S. Army Corps of Engineers (USACE) to create a generic format for electronic delivery of analytical data for environmental programs.

SEDD is being implemented in stages to allow analytical laboratories time to adapt their procedure to the new format.

- Stage 1 includes basic analytical data elements (such as the sample ID, analyte, result and qualifier) to convey results to the end user.
- Stage 2a adds method Quality Control (QC) data to Stage 1.
- Stage 2b adds instrument QC data to Stage 2a.
- Stage 3 adds additional measurement data to Stage 2b to allow for independent recalculation of reported results.
- Stage 4 adds the raw instrument data files to Stage 3.

The format is based on the eXtensible Markup Language (XML), which is rapidly becoming the standard for data transfer on the web. According to Microsoft, “XML is revolutionizing how applications talk to other applications – or more broadly, how computers talk to other computers – by providing an universal data format that lets data be easily adapted or transformed”. XML is not a programming language but rather is a format for structuring data. An XML file contains data nodes, elements and tags that describe the data in the file. In this way the file is self-descriptive and can be understood by any XML-compliant web application. An XML can also have a related Document Type Definitions (DTDs) file to impose a data format and allow for data verification.

SEDD and Web Services

Because SEDD uses XML, it is a perfect fit for use with a new Internet technology called Web Services. A web service can best be described as a “site intended for use by computer programs instead of by human beings” (Microsoft, 2002). Each web service is a small task-oriented application accessible through the Internet. Since web services use XML for transferring information, they can act as the bridge between different applications, computers, intranets and database systems. Again from Microsoft, web services “let applications share data, and - more powerfully - invoke capabilities from other applications, regardless of platform”.

The Internet Evolution

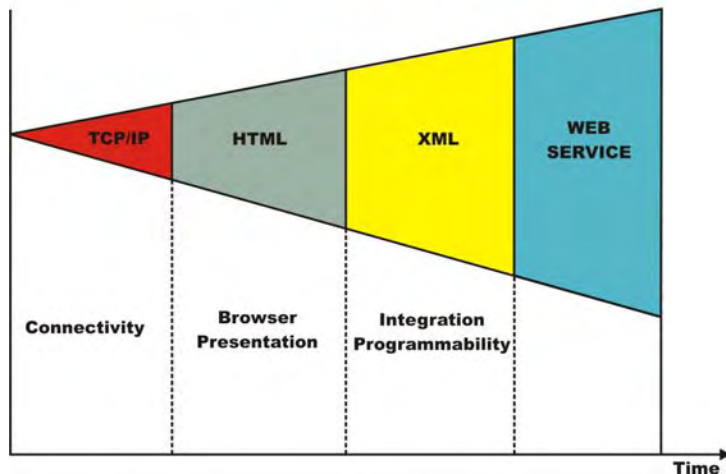


Figure 2. The Advent of Web Services

Common examples of web services include the various news and stock tickers, weather forecasts and sports score applications that can be installed on a desktop PC. The applications query a dedicated web service for the latest news, sports or weather in XML format and push the data on the desktop. Locus has developed a similar application that lets a user load a SEDD file into EIM.

The Locus SEDD 2a File Helper

Locus has developed the Locus SEDD 2a File Helper application to let the user:

- check a SEDD file with a DTD.
- load a SEDD file into EIM.



Figure 3. The Locus SEDD 2a File Helper

The application, which runs on a client PC, was developed using the Microsoft Visual Basic.NET programming language. Currently the application only supports Stage 2a SEDD v 5.0 files. The application calls a Visual Basic.NET Web service to perform the data transfer to EIM.

Checking a SEDD File

The “Check a SEDD File” button opens a form where a SEDD file can be selected.

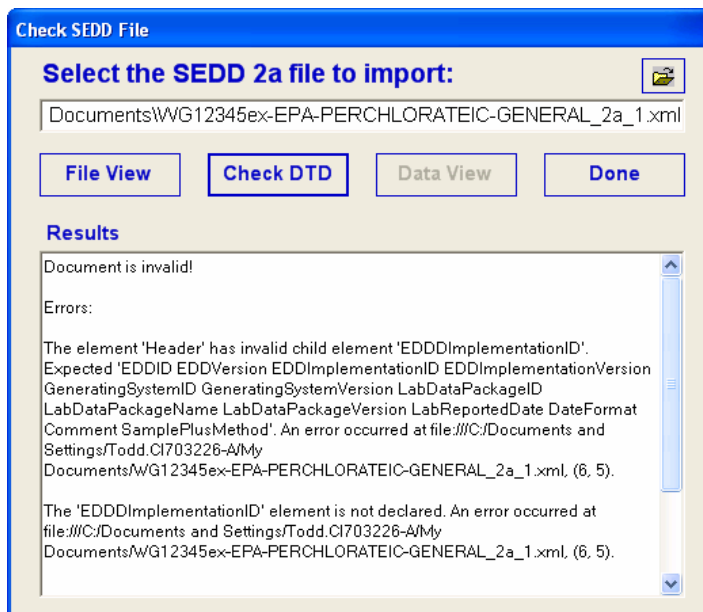


Figure 4. A SEDD File with DTD Errors

The user can preview the SEDD file before performing the file check with the “File View” button. The “Check DTD” button checks the file against the DTD referenced in the SEDD file. These DTDs are stored and updated on the Locus Web server. DTDs can be custom-built for different clients and programs.

The file checking is done on the local PC using Microsoft .NET XML validation objects. If a file passes the DTD check, the “Data View” button lets the user preview the file. If the file fails the check, the user is shown diagnostic information.

Loading a SEDD File to EIM

The “Load a SEDD File to EIM” button opens a form where a user logs in to EIM. The user must first select the target database and site in EIM. The “Load to EIM” button then transfers the file to a Locus web service. The web service loads the data into EIM and converts the SEDD data to EIM format using SQL stored procedures.

SEDD to EIM Table Translation

The SEDD stage 2a file has five nodes that are mapped to EIM tables.

- SEDD Header Node → EIM Dataset table.

- SEDD SamplePlusMethod Node → EIM Field Sample and Lab Sample tables.
- SEDD Reported Result Node → EIM Field Sample Result and Lab Sample Result tables.
- SEDD Analysis Node → EIM Field Sample Result and Lab Sample Result tables.
- SEDD PreparationPlusCleanup Node → EIM Field Sample Result and Lab Sample Result tables.

The SEDD Handling, AnalysisGroup and Analytes nodes are not yet used by Locus clients.

SEDD to EIM Field Translation

Converting the SEDD data elements to EIM fields is mostly a straightforward translation, with some issues.

- Some EIM fields are shorter than SEDD elements and must be checked for truncation; the DTD file can store these length requirements.
- The SEDD date/time elements must be split into separate date and time fields.
- The SEDD elements ResultBasis and MatrixID must be examined to set the EIM leached and filtered flag fields.
- Care must be taken to correctly handle the QC sample links in SEDD.

SEDD to EIM Value and Null Translation

In most cases the actual data in a SEDD element can be copied straight to the matching EIM field, but some valid value translations must be made. For example, a SEDD AnalyteType value of 'Surrogate' becomes an EIM value of 'SUR'. Also, some SEDD elements which allow NULL values must be handled separately. For example, a null AnalyteType in SEDD becomes a EIM value of 'TRG'.

EIM Error Checking

After the SEDD file has been translated to a set of EIM tables and field values, EIM runs checks on the data and reports errors caused by missing values, wrong data types and invalid sample relationships. The "View Error Summary" button shows a list of errors by type and column. The "View Errors" button shows all SEDD records with errors, linked to an error list for each record.

| Error No | Column | Error Count | Error |
|----------|------------------------|-------------|--|
| 9 | (null) | 4 | Duplicates Found In File |
| 2 | ANALYTICAL_METHOD | 329 | Entry Not In List Of Valid Values |
| 2 | FIELD_SAMPLE_ID | 189 | Entry Not In List Of Valid Values |
| 2 | LAB_ID | 329 | Entry Not In List Of Valid Values |
| 1 | LAB_RESULT | 2 | Required Value Is Missing |
| 1 | LAB_UNITS | 2 | Required Value Is Missing |
| 3 | LOWER_LIMIT | 210 | Numeric Value Is Required |
| 2 | PARAMETER_CODE | 290 | Entry Not In List Of Valid Values |
| 1 | REPORT_DETECTION_LIMIT | 189 | Required Value Is Missing |
| 14 | REPORT_RESULT | 189 | Required Value Calculated Or Derived From Data |
| 14 | REPORT_UNITS | 189 | Required Value Calculated Or Derived From Data |
| 3 | RPD | 255 | Numeric Value Is Required |
| 3 | SPIKE_RECOVERY | 219 | Numeric Value Is Required |
| 3 | UPPER_LIMIT | 210 | Numeric Value Is Required |

Figure 5. An Error Summary for EIM Errors

If a file has EIM errors, the user can login to the main EIM interface to review and fix errors and save the SEDD data to the final database tables. If a file has no EIM errors, the user must still login to EIM to sign off on the data and save it to the final tables.

Using SEDD Data in EIM

A laboratory could generate a SEDD file and upload it to the EIM system for data format verification by an engineer or geologist, and data validation by a chemist or data validator. Once the data is fixed and saved to EIM, the data can be queried into tables, spreadsheets or reports, posted to a map in a geographic information system (GIS) or exported back to SEDD or other formats for transmittal to regulatory agencies or other consultants or PRPs. The data can also be accessed from a user's PC using the Locus EIM Query web service. This web service lets users download EIM data directly into ESRI's ArcView9, Microsoft's Excel or AutoCAD's Autodesk.

Conclusions

- The SEDD format is based on XML and, hence, is well-suited for data transfer to databases using Web services.
- Some work is required to convert SEDD data to a specific format, but the work is made easier by the use of SQL stored procedures.
- Using DTDs with SEDD files enables SEDD file generators to ensure the files are consistent and correct before sending the files to the data loaders and users.
- A Web service for loading SEDD files can facilitate quick transfer of SEDD data into a data management application for use in queries, tables and maps as well as exports to other formats.
- SEDD helps make the analytical data "life cycle" quicker and more efficient from sample collection through lab analysis to final report preparation.

Web Services - Some Definitions

- SOAP (Simple Object Access Protocol) is the XML-based set of rules for the call-and-response communication between Web service-enabled applications. SOAP

is the glue that holds Web services together by ensuring reliable delivery of messages.

- WSDL (Web Services Description Language) describes the design of a Web Service so a client can discover how to invoke and use it.
- UDDI (Universal Description, Discovery, and Integration) is the standard for registering all available Web services in use. UDDI is like a phone book for locating a particular Web service.
- XML (eXtensible Markup Language) is rapidly becoming the de facto standard for transferring data between databases and applications on the World Wide Web.

For More Information

- World Wide Web Consortium: <http://www.w3.org>
- EPA: <http://www.epa.gov/cdx>
- SEDD: <http://www.epa.gov/superfund/programs/clp/sedd.htm>
- XML: <http://www.xml.org>
- Microsoft: <http://www.microsoft.com/net/basics/webservices.asp>
- UDDI: <http://www.uddi.org>, <http://services.xmethods.net/>
- Locus and EIM: <http://www.locustec.com>

References

The SEDD web site above was the primary source of information on the SEDD format. The Microsoft quotes were obtained from the above Microsoft link on web services. The World Wide Web Consortium web site above was the source of information on the XML format.