THE PETROLEUM INDUSTRY AND XML

The purpose of this White Paper is to provide accurate and balanced information about *eXtensible Markup Language (XML)* in support of the mission of the Petroleum Industry Data Exchange (PIDX) Standards Subcommittee (PSS) XML Task Group. The mission of this task group is:

> ***To establish, maintain and publish guidelines and resources for development and implementation of open and interoperable XML for use in the Petroleum Industry.***

The ultimate objective of this document is to assist the petroleum community with understanding this new technology and understanding how it may have an impact on industry business practices. This is a compilation prepared by business and technical experts in the petroleum community.

PIDX INTERNATIONAL
The place to go for
oil and gas e-business standards

PIDX EC Evolution: Past and Present

The Petroleum Industry Data Exchange, Inc. (PIDX) mission is to (1) influence the architecture and facilitate the implementation of effective standards and processes for electronic business communications within the petroleum industry community and (2) leverage the inherent value in existing EDI standards as well as actively pursue the benefits of the emerging e-Business technologies. Electronic Data Interchange (EDI) is the exchange of business information between companies using standard formats without the need for manual intervention. Electronic commerce made possible by EDI is rapidly gaining acceptance as a means of reducing costs and improving business-operating efficiencies in many industries.

EDI is both a *technique* and a *technology*. As a technique, electronic data interchange is the business of reliably exchanging data between independent computing systems, such as between a vendor and customers. As a technology, EDI is the means of formatting and transmitting that data. Together, these are *traditional* EDI.

EDI accomplishes the transmission of data between trading partners from computer application to computer application. Before EDI, data was transmitted using proprietary flat data file formats between trading partners. Individual companies had their own format for sending data files, requiring the receiver of the data to support different formats for each trading partner.

The American National Standards Institute Accredited Standards Committee (ANSI ASC) X12 was established more than 20 years ago to address this problem. It was tasked to develop uniform formats for transmitting purchase orders, invoices, and other common business documents to enable the computer systems of buyers and sellers to exchange pre-defined data over a secure, reliable network. This provided fast and accurate data that was otherwise subject to processing and handling delays inherent in surface mail, faxes and data entry.

EDI works by encoding data in a very specific, standard format. EDI communication consists of one or more *transaction sets* also known as *messages*, which is the equivalent of a business document such as a purchase order or an invoice. Each transaction set or message conforms to a subset of the standard. Within each message, delimiters define a series of *segments.* The first two or three characters in a line of data contain the segment tag or identifier. Within each segment, a series of codes, literal text data and numbers (separated by another delimiter) define the data being transmitted. The EDI-defined flat file is then sent from the sender to the receiver through either a direct model connection, via a private network connection or through a *Value-added Network (VAN).* Costs for using a VAN are based on frequency and volume, and the VAN supplies additional security and tracking functions. Recently, transmission via Internet has become possible.

The use of EDI is possible only because of the agreement between sender and receiver on the use of codes imbedded in each message. These agreements are captured in a *transaction set*. Initially, these formats and codes were developed ad hoc, with standards being introduced by national (e.g., ASC X12 in the USA) and international bodies (e.g., UN/CEFACT) beginning in the 1980s. These bodies have continued to maintain existing and develop new transaction sets as needed. There are also several industry standards that have been developed to address petroleum specific needs. (i.e., CDEX, RECON, TABS, etc.)

EDI is used extensively today by many industries to exchange electronic business documents.   Industry has invested billions of dollars into creating their own EDI infrastructures. EDI provides a highly structured organized set of rules based on industry standards (ASC X12 / UNEDIFACT).   These industry EDI standards allow for highly reliable data exchanges between partners.   EDI is continuing to grow, although the rate of EDI adoption is decreasing. EDI's market position has clearly not penetrated to the small and medium size companies.

XML is a language that has the potential to reach those electronic trading partners that find traditional EDI too challenging. EDI has revolutionized the way large organizations do business. It facilitates and promotes the exchange of fixed structured information between automated systems using pre-defined standard formats in both batch and real-time.

What is XML?

Extensible Markup Language (XML) is a World Wide Web Consortium (W3C) Recommendation that defines data meaning rather than its presentation. XML is a subset dialect of the Standard Generalized Markup Language (SGML), which was developed to interchange technical documentation and other forms of publishable data.

XML works by allowing users to define a hierarchical set of tags that are embedded into a file that contains the information being communicated. The tags (with starting and ending forms) explain exactly what the data in the tagged section of the document is intended to mean.

Each XML document may explicitly declare the source of the tag set that it employs, thus providing extremely valuable information to the reader of the file. Each set of tags is defined in a separate (usually web accessible) file, presently in the form of a *document type definition* (DTD). Because DTDs are very limited in their definitions of data types, an additional W3C specification for *XML Schema* will replace DTDs as the means to define XML tags. XML documents today generally contain data for a particular domain, e.g., a purchase order, a well log, a production report, etc. These are application level specifications for various kinds of data. In the near future, however, many XML documents will begin incorporating additional tags that establish a processing framework for the domain information within the document. These tag sets can be thought of as interoperability level specifications. There are several active initiatives that are specifying frameworks for electronic data interchange with XML, e.g., BizTalk, RosettaNet and ebXML. The W3C has produced a white paper comparing a number of XML protocols that fall into this category. To date, however, none of these initiatives dominate the commercial marketplace.

In addition to these definitional components, methods for applications to interact with XML are defined with the Document Object Model (DOM) and the Simple PIDX for XML (SAX). Another significant tool is W3C's XSL Transformation (XSLT), which allows XML messages to be sub-setted, reordered and converted into other forms using reusable, recursive templates. This allows XML to be converted to HTML for conventional web browsing, or into application specific structures for data input.

PIDX INTERNATIONAL
The place to go for
oil and gas e-business standards

The Impact of XML

Since XML is based upon SGML, it has a great deal of capability when dealing with human readable text. XML allows the organization of information to more closely reflect the forms that human readers understand while retaining a capability for machine processing. This provides an efficient structure for evaluating content for information that has heretofore been unmanageable prose.

As data is moved between formats (i.e., observation → paper → file → database), it is often altered to fit the requirements of the new format. These alterations often degrade the original meaning suggesting that the frequency and degree of change should be minimized whenever possible.

XML provides a methodology to describe the structure in which data resides (context), and these XML structures can be widely implemented, e.g., in sensors (observation), documents (paper), messages (file), applications (memory) and databases (disk storage). While not claiming universal applicability, XML-based data structures are significantly more resistant to the 'data rot' (decrease in value due to loss of context) that naturally occurs in most business processes.

The impact of XML on data is illustrated in the this diagram:
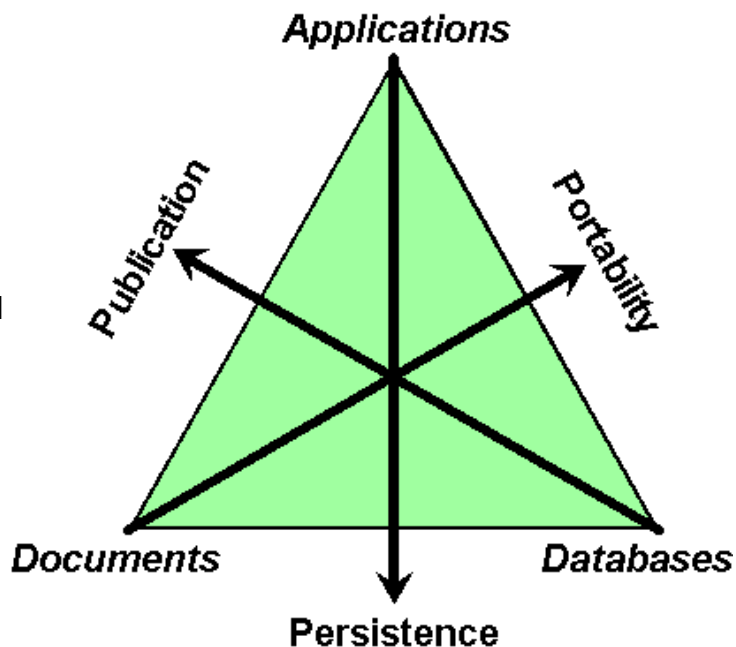
- **Data** is commonly distributed between applications, databases and documents, each with unique shortcomings

- **Persistence** for data in applications is improved by allowing data to be moved out of applications and into both documents and databases more easily and with less semantic loss

- **Portability** for data in documents is improved by making data in documents more compatible with the more structured form needed for input by applications and databases

- **Publication** for data in databases is improved by making database contents easier to locate, read and write through reducing the cost (and losses) of converting data into and out of a highly structured form

XML PROS AND CONS

PROS OF XML	CONS OF XML
• Human and machine readable • Self describing • Free and inexpensive tools • Future of application development • Real-time messaging tool • Complement to EDI • New format to exchange data • More consistent usage of data in documents, applications and databases (No Format Changes Required) • Making data used more consistently; content to create objects • Used by all industries • Lower transaction cost than VAN • Expanding trading community • Easier development • Reduction of VAN usage • Grow usage and expand data capture	• Lacks data standards and usage conventions • Files are much larger • Time for implementation • Learning curve and training needed • Data types not validated • Not a total technical solution • It is a language—not a process • New technology expense, investment required • Security, authentication, non-repudiation not inherent or part of the solution • Lack of standard application interfaces • Different infrastructure required

PIDX INTERNATIONAL
The place to go for
oil and gas e-business standards

Adopting XML

Within a few years, XML and eBusiness frameworks built on XML may create a new form of EDI. Traditional EDI is not endangered, but it will become a more specialized technology as XML-based EDI (XML/EDI) will gain acceptance for activities that involve human interaction or limited capital. XML has facilitated redefining processes and transactions used for eCommerce. XML is a new opportunity for companies to build on their EDI successes by utilizing this technology to increase their trading partner community, reduce total system costs, and provide flexibility with data content based on a company's business need.

Each company's transition from EDI to XML will occur at different rates depending upon current EDI commitments. Most businesses not currently using EDI will begin to use XML once:
1. the standards are well established,
2. the commercial tools are widely available, and
3. It is supported by either a significant business partner or as a vendor service.

Those currently using EDI will have several options. These options will be influenced by both internal (e.g., cost savings, higher degree of human interaction needed, connecting with new partners) and external factors (e.g., changes by business partners, new legislation). The spectrum of possible change includes:
1. Make no changes because (a) VAN (Value Added Network) security and reliability cannot be compromised, (b) EDI formats are very stable, and (c) all business partners are established and remaining on VAN-based EDI;
2. Switch from VAN-based to Internet-based EDI, continuing to use the traditional EDI format;
3. Encapsulate the traditional EDI format within a wrapper of XML tags, with the degree of replacement gradually increasing with time; or
4. Install a parallel XML-based system that will eventually replace the traditional EDI system.
5. Install new applications that replace current functionality.

As a result of EDI's high entry cost, approximately 2% of the world's businesses use EDI today. XML has the potential to increase usage toward the 70-80% range, as the cost of entry becomes less of a barrier to small and medium sized enterprises thereby increasing trading partner participation. XML also allows businesses added flexibility with data handling.

Assuming that XML will be based upon the popular adoption of formal XML standards, then two or more years may be required for the formal adoption of XML standards.

The PIDX Direction on XML

PIDX recognizes that XML has the potential to significantly expand the number of trading partners and reduce the costs of electronic commerce. Further, PIDX believes that XML will complement rather than conflict with conventional electronic data interchange (EDI).

PIDX intends to work towards the standardization of XML, the harmonization of XML with existing EDI standards, and the development of XML frameworks to make XML simpler to develop, maintain and use. It is recognized that XML standards are still emerging. PSS will assess the viability and make recommendations on the direction as the technology matures.

PIDX is committed to working with other EDI and XML initiatives and other standards bodies to ensure there are open, interoperable XML standards. The initial focus of PIDX will be data standardization specific for our industry and core component definitions as business needs are identified. The PIDX Standards Subcommittee participates in XML development as follows:

- Member of the BizTalk Steering Committee
- ebXML working group participation and development through ASCX12
- XML pilot development with user and technical participation
- Monitor other Industry and XML standards emerging, i.e. RosettaNet, OAGIS, etc.
- Expand the Petroleum Industry Data Dictionary(PIDD) for use in defining XML tags
- Develop PSS XML Review Process and Developers Guidelines
- Develop Recommended Practices for XML Documents

PIDX INTERNATIONAL
The place to go for
oil and gas e-business standards

PIDX Resources

PIDD

The Petroleum Industry has developed a common Petroleum Industry Data Dictionary (PIDD) which defines terms unique to our industry. PIDX is expanding this dictionary to include XML tag names, attributes and syntax. As these terms have been used in EDI, it is intended that these XML tags will be used in XML.

The addition of XML tags to PIDD is initially being constructed within the framework proposed by the BizCodes initiative (http://www.bizcodes.org/) to the ebXML project (http://www.ebxml.org/).

The maintenance of the PIDD will continue as defined by the PIDX Standards Subcommittee. New terms and modifications to existing terms may be submitted by email. New terms must include definitions. Existing terms will not be removed but corrections may be offered. The PIDX Standards Subcommittee will review all XML tags before being approved for the PIDD.

Training

PIDX will develop and provide training classes on XML suited to the petroleum industry. This training will range from half-day tutorials to multi-day classes. The purpose of this effort is to introduce XML as a technology of increasing importance to E&P professionals, and to widen and strengthen the commitment to standardization within our industry.

GLOSSARY OF TERMS

AFEX (Authorization for Expenditures)
PIDX application supporting the AFE balloting and approval process between operators and working interest owners. Part of the PIDX Joint Interest Data Exchange User Group.

ANSI ASC X12 (American National Standards Institute Accredited Standards Committee X12)
X12 is an Accredited Standards Committee operating under the procedures of the American National Standards Institute. ASC X12 develops, maintains, interprets, publishes and promotes the proper use of Electronic Data Interchange standards.

Auto-Fax
A system which allows a user to request on-line documents which are then automatically sent to the requestor via a specified facsimile number.

AVNET (Electronic Aviation Fuel Sales)
PIDX application designed to enable the transmission and automated processing of aviation fuel, into-plane, and bulk sale related documents between Trading Partners.

Bar Coding
A symbology using a series of bars and spaces in which information can be encoded and then read by a scanning device communicating product and/or transaction data.

CDEX (Check Stub & Royalty Payments Data Exchange)
PIDX application designed to provide formatted data on crude oil and gas lease check stub detail and enables the direct transmission of royalty payment and remittance information from lease operators to owners.

CODE (Crude Oil Data Exchange)
PIDX application for the exchange of crude oil and condensate run tickets, statements and tank increments between producers and transporters.

DES (Data Encryption Standard)
Data Encryption Standard (DES) is a widely-used method of data encryption using a private (secret) key that was judged so difficult to break by the U.S. government that it was restricted for exportation to other countries. There are 72,000,000,000,000,000 (72 quadrillion) or more possible encryption keys that can be used. For each given message, the key is chosen at random from among this enormous number of keys. Like other private key cryptographic methods, both the sender and the receiver must know and use the same private key.

Digital Signature

An electronic rather than a written signature that can be used by someone to authenticate the identity of the sender of a message or of the signer of a document. It can also be used to ensure that the original content of the message or document that has been conveyed is unchanged. Additional benefits to the use of a digital signature are that it is easily transportable, cannot be easily repudiated, cannot be imitated by someone else, and can be automatically time-stamped.

EC (Electronic Commerce)

The integration of EDI, electronic mail, electronic funds transfer, imaging, and similar techniques into a comprehensive, electronic-based system for conducting business transactions throughout an enterprise.

EDI (Electronic Data Interchange)

EDI is the computer-to-computer exchange of machine-processable data in a standard format.

Electronic Catalog

Electronic mechanism for distributing product information between manufacturers / vendors and their customers.

E-mail

A system whereby a computer user can exchange messages with other computer users via a communications network.

ERP (Enterprise Resource Planning)

An industry term for the broad set of activities supported by multi-module application software that help a manufacturer or other business manage the important parts of its business, including product planning, parts purchasing, maintaining inventories, interacting with suppliers, providing customer service, and tracking orders. ERP can also include application modules for the finance and human resources aspects of a business.

Evaluated Receipts Settlement (ERS)

A business process in which payment is triggered by the receipt of goods rather than by receipt of an invoice.

Extranet

An Extranet is a private network that uses the Internet protocols and the public telecommunication system to securely share part of a business's information or operations with suppliers, vendors, partners, customers, or other businesses.

FTE (Full-time Equivalent)

Counts all employees based on the number of hours worked in a year. For example, a person that worked ten hours a week for one year or a person who worked forty hours a week for only three months would each be counted as one-fourth of a "Full-time Equivalent".

Global Positioning Technology

A technology that uses radio satellites for three-dimensional position, velocity and time information.

GRADE (Gas Revenue Accounting Data Exchange)
PIDX application designed to facilitate the transmission of natural gas liquids and natural gas metered and allocated volumes, test data and plant/lease settlement data.

HTML (Hypertext Markup Language)
The set of "markup" symbols or codes inserted in a file intended for display on a World Wide Web browser. The markup tells the Web browser how to display a Web page's words and images for the user. The individual markup codes are referred to as elements (but many people also refer to them as tags).

Imaging Technology
Technology which allows the alteration or manipulation of images that have been scanned or captured by a digital recording device.

Internet
The world's largest telecommunications inter-network, comprising large backbone nets and an array of regional and local networks worldwide.

Intranet
A private network that is contained within an enterprise. It may consist of many interlinked local area networks and also use leased lines in the wide area network. Typically, an Intranet includes connections through one or more gateway computers to the outside Internet.

JIBE (Joint Interest Billing Exchange)
PIDX application designed for the exchange of joint interest bills and related detail between operators and working interest owners. Part of the PIDX Joint Interest Data Exchange User Group.

JUPUG (Joint Utility/Petroleum User Group)
Joint PIDX/utility industry application that defines the data required for electronic invoicing and payment documents between petroleum, telecommunications, and electric utility companies.

Knowledge Management
The name of a relatively new concept in which an enterprise consciously and comprehensively gathers, organizes, shares, and analyzes its knowledge to further its aims.

MSDS (Material Safety Data Sheets)
PIDX application that defines the data required for the electronic transmission of Material Safety Data Sheet information used by petroleum and chemical companies.

PETNET (Petroleum Product Network)

PIDX application (using ASC X12 864 format) designed to help expedite the collection of accurate and consistent bill of lading and bulk custody transfer information on finished product inter-company transactions. Part of the PIDX Downstream EDI User Group.

PETROEX (Petroleum Product Exchange)
PIDX application (using PETRODEX format) designed to help expedite the collection of accurate and consistent bill of lading and bulk custody transfer information on finished product inter-company transactions. Part of the PIDX Downstream EDI User Group.

PGP (Pretty Good Privacy)
A popular program used to encrypt and decrypt e-mail over the Internet. It can also be used to send an encrypted digital signature that lets the receiver verify the sender's identity and know that the message was not changed en route.

PIDD (Petroleum Industry Data Dictionary group)
PIDD is the caretaker of the language of the petroleum industry. The role of the PIDD is to be the repository for the preferred meanings of business and technical terms used within the petroleum industry.

PIDX (Petroleum Industry Data Exchange, Inc.)
Formerly the Electronic Data Interchange Committee of the American Petroleum Institute. Since 2011, PIDX has operated as an independent 501(c) (60 trade association. It is also referred to as PIDX International.

PIPENET (Pipeline Information)
PIDX application designed for the transmission of refined product and crude oil pipeline data between pipeline companies, shippers, and refiners.

P&MM (Purchasing and Materials Management)
PIDX application providing informational requirements for purchasing, inventory, logistics and material control documents.

Portal

Generally synonymous with *gateway*, for a World Wide Web site that is or proposes to be a major starting site for users when they get connected to the Web or that users tend to visit as an anchor site. There are general portals and specialized or niche portals.

Purchasing/Procurement Card
A limited use, direct credit purchase card for use by a specific individual to facilitate the purchase of specific goods and services.

RECON (Exchange Reconciliation System)
PIDX application designed to provide an electronic means of obtaining information on unmatched exchange transactions on a contract and product basis. Part of the PIDX Downstream EDI User Group.

REGS (Regulatory Data Exchange)
PIDX application that serves as the focal point between the petroleum industry and federal/state regulatory agencies for upstream royalty, production and severance tax reporting.

RFG Reporting (Reformulated Gasoline)
EDI application designed for the electronic transmission of reformulated gasoline reports to the U.S. Environmental Protection Agency.

RSA (Rivest-Shamir-Adleman)
RSA is an Internet encryption and authentication system that uses an algorithm developed in 1977 by Ron Rivest, Adi Shamir, and Leonard Adleman. The RSA algorithm is the most commonly used encryption and authentication algorithm and is included as part of the Web browsers from Netscape and Microsoft. It's also part of Lotus Notes, Intuits Quicken, and many other products.

TABS (Terminal Administration & Billing System)
PIDX application designed to provide credit and product authorization at exchange terminals while capturing bill of lading information. Part of the PIDX Downstream EDI User Group.

TBWG (Telephone Bill Work Group)
EDI application designed to facilitate the electronic exchange of telephone invoices. Developed in conjunction with the Joint Utility/Petroleum User Group (JUPUG) and the Tele-Communications Industry Forum (TCIF).

TPA (Trading Partner Agreement)
A written instrument of understanding negotiated between EDI trading partners that specifies contractual matters and protocols of governing EDI transactions.

UNIX
UNIX is an operating system that originated at Bell Labs in 1969 as an interactive time-sharing system. Ken Thompson and Dennis Ritchie are considered the inventors of UNIX. The name (pronounced YEW-nihks) was a pun based on an earlier system, Multics. In 1974, UNIX became the first operating system written in the C language.

Voice Recognition
Voice or speech recognition is the ability of a machine or program to receive and interpret dictation, or to understand and carry out spoken commands.

Windows NT
Windows NT is the Microsoft Windows personal computer operating system designed for users and businesses needing advanced capability. Windows NT (which may originally have stood for "New Technology," although Microsoft doesn't say) is actually two products: Microsoft NT Workstation and Microsoft NT Server. The Workstation is designed for users,

especially business users, who need faster performance and a system a little more fail-safe than Windows 95 and Windows 98).

WITS (Wellsite Information Transfer System)
PIDX application that provides instructions and the format for transmitting real-time drilling parameters and information to joint interest partners.

WODEX (Well Operating Data Exchange)
PIDX application designed for the exchange of daily drilling and well operations reports between operators and working interest owners. Part of the PIDX Joint Interest Data Exchange User Group.

Workflow Technology
The automated coordination, control and communication of work, both people and computers, in the context of business processes, through the execution of software in a network of computers, whose order of execution is controlled by a computerized representation of the business processes.

XML (Extensible Markup Language)
A flexible way to create common information formats and share both the format and the data on the World Wide Web, Intranets, and elsewhere.