White Papers of Three Work Groups: PKI, Student Identifier, and XML

A Publication of the
Postsecondary Electronic Standards Council
Washington, DC
May, 2000

POSTSECONDARY ELECTRONIC STANDARDS COUNCIL

2000

Who we are…

The Postsecondary Electronic Standards Council is a membership organization open to any group with an interest in promoting or using electronic standards for data sharing within the higher education community. This community includes higher education associations; postsecondary institutions; data, software, and service providers; as well as state and federal government agencies.

Our diverse membership includes…

Associations

American Association of Collegiate Registrars and Admissions Officers
Coalition of Higher Education Assistance Organizations
EDUCAUSE
National Association of College and University Business Officers
National Association of Student Loan Administrators
National Association of Student Financial Aid Administrators

Postsecondary Education Institutions

Oregon Health Sciences University
The George Washington University
Mississippi State University
University of Illinois at Chicago
University of Iowa
University of Northern Iowa
University of Phoenix
University of Texas at Austin
Virginia Polytechnic Institute

Federal Agencies

U.S. Department of Education
U.S. Immigration and Naturalization Service

Software and Service Providers

ACT, Inc.
CDSI Educational Services, Inc.
Datatel, Inc.
Educational Testing Service
Embark.com
Harbinger
IMS Developers' Network
KPMG Consulting
Law School Admission Council
National Computer Systems
PeopleSoft
Peterson's Publishing Group
ScienceWise.com
SCT Corporation
The College Board

Lenders, Loan Guarantors and Servicers

Citibank
Sallie Mae Servicing Corporation
Student Loan Servicing Alliance
USA Group

What we've accomplished…

Work Groups

- Modification of existing EDI transaction sets to accommodate the sending of prospect information, Taxpayer Relief Act reporting data, and enrollment reporting data to insurance companies and the National Student Loan Clearinghouse
- Study of topics of interest to PESC membership: public key infrastructure (PKI); student identifiers; and Extensible Markup Language (XML).

Education/Training/Information Sharing

- Publication of *The Standard*, the monthly electronic newsletter of PESC; journal articles; and press releases
- Up-to-date web site with information on events, standards news, and useful resources
- Presentations at national meetings of NASFAA, AACRAO, NACAC, Society of Research Administrators, COHEAO, and ACUTA; regional meetings of The College Board and SACAC; meetings of the ACE Commission on Adult Learning and Educational Credentials and the Federal Demonstration Project; the EDI/EC in Education Conference and the HEWI Conference
- Host of annual Conference on Electronic Standards in Higher Education
- Publication of white papers on public key infrastructure (PKI), student identifiers, and Extensive Markup Language (XML).
- Representation on community task forces
- Best Practices Competition promoting the use of standards in higher education

Standards Support

- Development and publication of implementation guides for modified EDI transaction sets for education
- Maintenance of industry code set archive
- Representation and elected leadership in ANSI ASC X12 Education Administration
- Distribution of the Postsecondary Institution Crosswalk Table
- Representation in ebXML, an international effort to set specifications for XML
- Publication of conference papers on issues of standardization in education

1999-2000 Service Awards

David Stones, Chair, Student Identifier Work Group
Timothy Pavlick, Chair, XML Work Group
David Leonard, Chair, PKI Work Group

1999 Best Practices Awards

Ontario Universities' Application Centre

What we plan to do…

Work Groups

- Study of the creation of a PESC data definition repository for higher education
- Continuous survey of PKI activities in the higher education community
- Monitor and influence XML development to avoid overlapping and competing efforts
- Build on the recent study of existing student identifiers by investigating business needs, market forces, interested organizations, and technologies related to moving to a single student identifier

Education/Training/Information Sharing

- Articles to appear in industry publications
- Presentations at MOSIS and other education conferences
- Publication and distribution of conference presentations

Standards Support

- Steering Committee membership in the Modernization Partnership Forum
- Activity representation in work groups of the Modernization Partnership Forum
- Participation in X12 transition to process-centric focus, development of business process models, and creation of a repository for these models
- Data sharing initiative between schools and State Farm Insurance Company

POSTSECONDARY ELECTRONIC STANDARDS COUNCIL

One Dupont Circle, N.W.
Suite 520
Washington, D.C. 20036
(202) 293-7383
(202) 872-8857 FAX

1999-2000 STEERING COMMITTEE

C.J. Thoma
Chair
Student Loan Servicing Alliance

Jerald Bracken
Secretary/Treasurer
American Association of Collegiate Registrars and Admissions Officers

Bruce Bachman
Law School Admission Council

Lysbeth Bainbridge
Postsecondary Electronic Standards Council

Jackie Kessler
Systems and Computer Technology Corporation

A. Dallas Martin
National Association of Student Financial Aid Administrators

Mary Neary-Morley
PeopleSoft

Keith Riccitelli
USA Group

David Stones
Harbinger

Visit our web site

www.StandardsCouncil.org

for standards news, presentations, and white papers; technical resources; and information about our next membership year beginning July 1, 2000.

POSTSECONDARY ELECTRONIC STANDARDS COUNCIL

Membership Form
July 1, 2000 - June 30, 2001

Organization Name

Voting Member (or Official Contact) Name and Title

Street Address

City, State and Zip

Phone Fax E-mail Address

Membership Category

- ☐ Nonprofit association with revenues over $2,000,000 (annual dues $10,000)
- ☐ Nonprofit association with revenues under $2,000,000 (annual dues $7,500)
- ☐ Nonprofit association with revenues under $500,000 (annual dues $3,500)
- ☐ Nonprofit association with revenues under $200,000 (annual dues $2,000)
- ☐ State or Federal government agency (annual dues $10,000)
- ☐ Large commercial organization* (annual dues $10,000)
- ☐ Smaller commercial organization** (annual dues $7,500)
- ☐ Non-voting member (annual dues $2,000)
- ☐ Non-voting postsecondary educational institution (annual dues $500)

*Corporations and companies with gross revenue greater than $200,000,000; banks, S&Ls, and credit unions with gross assets greater than $1,000,000,000
**Corporations and companies with gross revenue less than $200,000,000; banks, S&Ls, and credit unions with gross assets less than $1,000,000,000

Areas of Interest

- ☐ Participate in work groups of interest to the organization
- ☐ Share expertise and information
- ☐ Gather information

Please complete the membership form and return to:

Betsy Bainbridge, Executive Director
Postsecondary Electronic Standards Council
One Dupont Circle, NW, Suite 520
Washington, DC 20036
fax: (202) 872-8857
email: BainbridgeL@aacrao.nche.edu

PKI Work Group

Public Key Infrastructure and Higher Education: An Introduction

A Publication of the
Postsecondary Electronic Standards Council
Washington, DC
May, 2000

Table of Contents

PKI Work Group Members

David Leonard, USA Group, Chair
John Barkley, NIST
Andrew Boots, US Department of Education
Michael Capps, Ed Fund
David Garver, ScienceWise
Dan Geller, NCS
Casey Lide, EDUCAUSE
Kelly Newcomb, TGSLC
Jan Nielsen, Campus Pipeline
Wally Reeves, University of Texas at Austin
Pat Salava, Educational Testing Service
Michael Sessa, NASLA
Paul Soohoo, Educational Testing Service
Shelby Stanfield, University of Texas at Austin
Craig Yamamoto, Ed Fund

This document was produced by members of the PKI Work Group in collaboration with technical consultants from Digital Signature Trust

Keren Cummins, Vice President, Government Services
Debbie Blanchard, Sales Engineer
Louis Jurgens, Sales Manager
Tim Pinegar, Senior Engineer

The Chair wishes to give special recognition to Work Group members Andy, Dan, Casey, Wally, and Paul and staff members Betsy and Andrew for their contributions.

~

The Postsecondary Electronic Standards Council
One Dupont Circle, NW, Suite 520
Washington, DC 20036
(202) 293-7383
http://www.StandardsCouncil.org

© May 2000

Public Key Infrastructure and Higher Education: An Introduction

Executive Summary

The Postsecondary Electronic Standards Council (PESC) created a Public Key Infrastructure (PKI) Work Group to determine how PKI can serve the needs of the higher education community. In this paper we examine the purpose of PKI and how it works in general, show how it has been implemented to date in higher education and elsewhere, discuss various models of PKI and the issues involved in an open PKI system, and present our recommendations for PESC support of PKI in higher education.

Introduction

PKI can provide assurances that data sent electronically over open networks arrive unaltered, unseen by unauthorized persons, from an authenticated sender who cannot deny sending the data. Without some or all of these assurances, there is not the required secure environment for many electronic transactions to take place.

As students, schools, lenders, service providers, and state and Federal agencies move from paper to electronic data exchanges of student data over the Internet, security concerns become important not only to the senders and receivers of this data but also to the public at large. Federal laws, regulations, and trading partner requirements for providing security with traditional paper processes and limited electronic exchanges must be adhered to in electronic data exchanges in an open environment. Digital signatures and encryption achieved through the use of a PKI can satisfy the required security attributes.

Discussion

A PKI, based on asymmetric or public key cryptography, incorporates the hardware, software, policies and procedures that allow previously unknown parties to exchange data in a secure environment. Two PKI models—closed (or enterprise) and network (or community of interest)—are found in use in various industries. A third model—open—is still in the conceptual stages. When fully realized, it will make available a "portable Internet credential" with broad utility even outside the originating community. The PESC community, with its singular focus on the entire higher education enterprise, is broad enough to feel the need for the open model PKI.

Individual universities and university systems are establishing their own closed PKIs to allow students, faculty, and administrators to use information technologies to improve the way they accomplish their business. Review of other network and open models such as the US General Services Administration ACES project and the American Bankers Association TrustID initiative, provide examples for higher education to consider in the move from institutional closed PKIs to an eventual network and open PKI.

Challenges for implementation of a broader-based PKI in higher education revolve around acceptance of the technology by users, trust in the process, interoperability, the issues surrounding provision of certification authority services, and costs.

Recommendations

The PESC PKI Work Group recommends that PESC encourage electronic commerce in higher education by fully supporting and promoting the use of PKI. Its role would focus on research and sharing of technical, legal and policy information with members of the higher education community and on the cooperation with existing and new initiatives to employ PKI technology in higher education. The Work Group recommends a standing committee be formed and a PKI expert be engaged to provide appropriate information on the topic on a continuing basis.

1 Introduction

The revolution in the delivery of services through electronic means that is taking place in the commercial and government worlds has a special resonance for the higher education community -- a community that exists entirely for the communication of knowledge and information. The Postsecondary Electronic Standards Council (PESC) comprises a diverse group of organizations that have entered into partnership within the higher education community for the purpose of improving service and controlling costs through the promotion of standards for data sharing.

The higher education community and those that support it--campus administrative offices, testing services, financial aid lenders, servicers and guarantors, and state and federal agencies--are finding more ways to provide better service and cut costs by taking advantage of electronic delivery of computerized data. As paper processes are replaced by electronic data sharing, current and potential trading partners must agree on data formats, code sets, encryption schema, and other standards to facilitate the flow of information. To that end, the PESC Public Key Infrastructure (PKI) Work Group was charged to investigate industry, state, and federal efforts to establish and regulate public key infrastructure (PKI) components of public key cryptography, digital signatures, digital certificates and certification authorities for secure Internet data exchanges.

This paper is intended to examine, from the perspective of the PESC, how PKI, or public key infrastructure, can provide a cornerstone for furthering that revolution in the higher education community. The paper will examine the purpose of PKI and how PKI works in general; how PKI is working in current implementations in higher education and elsewhere; what special concerns higher education may need to address in the future; major issues in the implementation of PKI in an open system, and the Work Group's recommendations for further action by PESC.

1.1 The Purpose of Public Key Infrastructure (PKI)

In the conventional world of business and professional communications, basic assumptions can be made when a transaction is conducted using original signed documents, delivered in sealed packages directly to the participating parties. Generally, organizations that use the public network to transmit business communications must understand with whom they are dealing; that no one has altered the content of the communication after it was signed; that no one unauthorized has seen the information; and that the signer of the document cannot deny having sent the document. These are the fundamental assurances that make transactions between remote parties possible.

Simply stated, the most fundamental purpose of a PKI – a public key infrastructure – is to provide these same types of assurances about the content and parties to transactions that take place over an open network, sometimes among parties not otherwise known to each other. By enabling the use of digital signatures and encryption, PKI can provide the same four basic security services for today's data transmissions:

- *Authentication* — Ensure that transmissions and messages, and their originators, are authentic, and that a recipient is eligible to receive specific categories of information.
- *Data Integrity* — Ensure that data is unchanged from its source and has not been accidentally or maliciously altered.
- *Non-repudiation* — Ensure strong and substantial evidence is available to the sender of data that the data has been delivered (with the cooperation of the recipient), and, to the recipient, of the sender's identity, sufficient to prevent either from successfully denying having sent or received the data. This includes the ability of a third party to verify the integrity and origin of the data.
- *Confidentiality* — Ensure that information can be read only by authorized entities.

In addition, PKI can also offer additional security services not offered by a paper process, including:

- *Authorization* –- Ensure that the user, either sender or recipient, is authorized for access to data, systems, or applications
- *Availability* – Ensure that legitimate users are not unduly denied access to information and resources

The term “public key infrastructure” or “PKI” refers to a complex suite of hardware, software and particular cryptographic components, combined with adherence to policies and procedures that enable business applications to operate in a secure environment. The particular cryptographic components used are those of public key, or asymmetric, cryptography. To discuss how PKI can serve higher education it is fruitful to review how those technologies, particularly digital signatures, actually work.

1.2 Understanding Public Key Infrastructure and Technology

To understand a PKI, it is important to understand how public key technology is used to encrypt and sign an electronic message. A digital signature is not a digitized image of a handwritten signature, nor is it a PIN or a password. Rather, a digital signature is an attachment to an electronic message that includes a mathematical digest of the message, created using public key cryptography. As a result, a digital signature is specific both to the signer of an electronic document or message as well as to the electronic document or message itself. Thus, a digital signature can be used to affirmatively identify both.

Cryptography Overview

There are two fundamental types of cryptography, symmetric cryptography and asymmetric or public key cryptography. Each of these cryptographic systems has distinct characteristics and is used in different ways to provide general security services.

Symmetric cryptography is the most familiar. It is based on a shared secret, or key, and works well within isolated environments. A simple example of symmetric cryptography is seen when children create a "secret code" by converting all the letters of the alphabet to their numeric equivalents, 1-26, and then adding, say, the number five. The child receiving the secret message has only to subtract five from each number in order to be able to reconstitute the correct numbers for the original message. The shared secret is the number five. In real commercial uses of symmetric cryptography, the algorithm used to encrypt may be enormously more complex. The challenge is communicating a secret key between the sender and receiver without anyone else finding out, because anyone who might intercept the key would be able to read, modify, or forge messages that were encrypted or authenticated using that key.

The inherent problem with symmetric cryptography is one of scalability. In order for the communications to be confidential, the exchange of a key, or shared secret, must be done securely between every pair of participants. Obviously, this type of secure distribution becomes increasingly difficult as the number of different people with whom you want to communicate securely grows.

The other encryption technique is asymmetric or public key cryptography which involves an asymmetric key pair. This key pair comprises what is referred to as a public key and a private key. The public key, as its name suggests, may be freely disseminated. This key does not need to be kept confidential. The private key, on the other hand, must be kept secret. The owner of the key pair must guard his private key closely, as sender authenticity and non-repudiation are based on the signer having sole access to his private key. Furthermore, there is no longer a need for the parties to a transaction to exchange any secret information.

There are several important characteristics of these key pairs. First, while they are mathematically related to each other, it is impossible to calculate one key from the other. Therefore, the private key cannot be compromised through knowledge of the associated public key. Second, each key in the key pair performs the inverse function of the other. What one key does, only the other can undo.

Digital Signatures and Public Key Encryption

Digital signatures are created using asymmetric or public key cryptography. The sender can produce a one-way hash of an entire message or a synopsis of the message. The sender encrypts the hash with their *private* key, thereby signing the document. When the message is received, the receiver recreates the message's hash from the message he received using the senders *public* key. If this process produces an identical result, then the recipient knows the following:

- that the person who sent the message is the holder of the private key associated with the public key used to verify the message;
- that that person cannot deny having sent the message;
- that the message has not been altered or modified in transit

If the sender also wishes to encrypt the message so that only the receiver can read it, instead of using his own key he must retrieve the public key of the intended receiver. Encryption conducted with that key is one-way, and only the private key of the intended receiver can be used to decrypt the message. Luckily, most PKI-enabled applications (e.g., S/MIME email) perform all of this "cryptomagic" for the user automatically.

Where Does *Infrastructure* Come In?

While the public key technology described to this point is available and in use, there are still shortcomings – at least from the standpoint of secure transactions among strangers. As seen from the example above, in order to validate a digitally signed document, the recipient must have access to the signer's public key. Likewise, in order to encrypt a message to a specific recipient the sender must have access to that recipient's public key. How is this obtained?

The signer may provide it directly, but there are two problems with this method. First, it is not a scalable solution. Taking into account the millions of Internet transactions among relative strangers, it is not realistic for all users to mail their public keys to everyone with whom they want to do business. Second, there is the risk of identity fraud. If the signer sends you a public key claiming to belong to John Doe, how do you know that it really was issued to John Doe?

The question arises: Who will attest that a particular public key really belongs to a specified individual? The use of public key technology creates the need for an entity to serve as a *trusted third party* (TTP) to vouch for individuals' identities and their relationship to their public keys. This entity, in public key infrastructure terminology, is referred to as a certification authority, or CA.

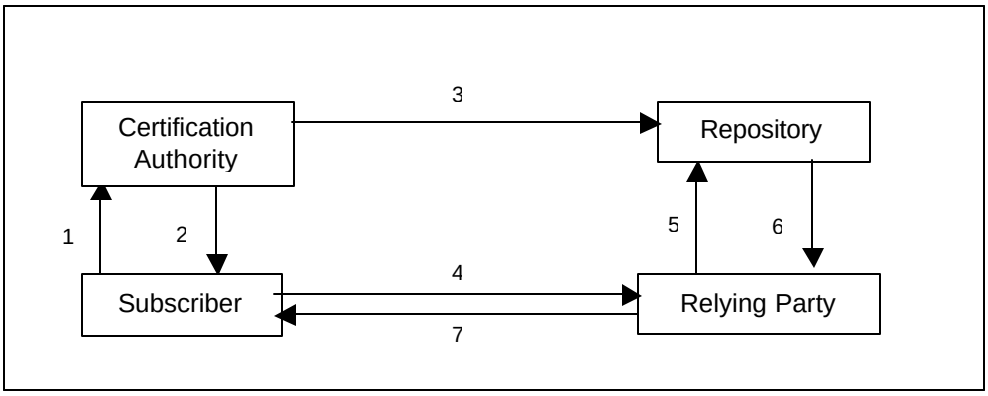
The CA is a trusted third party that issues digital certificates to its subscribers. The digital certificate is the document that binds a person's (or entity's) identity to the key pair used to digitally sign electronic communications. Digital certificates contain the name of the subscriber, the subscriber's public key, the issuing CA's public key, and possibly other pertinent information about the subscriber. The digital certificate is signed by the issuing CA, so that the information in the certificate cannot be altered. They can be revoked if the private key becomes compromised, or if there is some other change to the accuracy of the identity information, such as separation from an organization.

Certificates are typically stored in an on-line, publicly accessible repository. In addition to valid certificates, the repository also maintains an up-to-date listing of all the unexpired certificates which have been revoked.

What might these relationships look like over time? On a periodic basis, perhaps annually, a subscriber applies to a certificate authority (CA) for a digital certificate. The CA verifies the identity of the subscriber and issues the digital certificate. The CA publishes the certificate to a repository. Then for each transaction, the subscriber can now use the private key of the digital certificate to sign electronic messages.

Relying parties can now receive the message, verify the digital signature with the public key of the subscriber, and check the repository for validity and status of the digital certificate of the subscriber. The following figure, *PKI Process Flow*, highlights this process.

PKI Process Flow



Step 1. Subscriber applies to Certification Authority for Digital Certificate.
Step 2. CA verifies identity of Subscriber and issues Digital Certificate.
Step 3. CA publishes Certificate to Repository.
Step 4. Subscriber digitally signs electronic message with Private Key to ensure Sender Authenticity, Message Integrity and Non-Repudiation and sends to Relying Party.
Step 5. Relying Party receives message, verifies Digital Signature with Subscriber's Public Key, and goes to Repository to check status and validity of Subscriber's Certificate.
Step 6. Repository returns results of status check on Subscriber's Certificate to Relying Party.
Step 7. The Relying Party now can make an informed decision as to whether or not to trust the subscriber's message.

Properly configured, today's browsers and other PKI-enabled clients make these steps practically seamless to the participants involved.

The Repository plays a similar role in encryption of an electronic message so that the confidentiality of the message content is maintained. The subscriber encrypts the message with the Relying Party's public key. Only the Relying Party can decrypt the message with his private key.

1.3 Policy – The PKI Document Set

Published documentation for a certificate authority articulates the measures taken by that certificate authority to authenticate certificate subjects, such as users, servers, and organizations, and to protect the certificate authority. The minimal set of documents to be prepared by the certificate authority is the Concept of Operations (ConOps), Certificate Policy (CP), and Certificate Practices Statement (CPS). The CP and the CPS are typically published and are essential for any certification authority to define the technical, procedural, and legal foundations contributing to the trustworthiness of the certificate authority. To ensure the integrity of a PKI environment, a process must be established to verify compliance with these policies and practices. Each of these documents is described in the following sections.

Concept of Operations

The Con Ops is a high level document describing policies and basic concepts of operation unique to the participating organization's requirements and how the application of PKI technology would work. A Con Ops document includes a survey of existing business practices and highlights areas where change may be

needed to accommodate digital signature implementation. Additionally, the Con Ops would include an overview of the proposed system design and architecture.

Certificate Policy

The CP provides the fundamental grounding and basis for all PKI-related implementations, including applications. It is a detailed description of the operating rules surrounding the implementation of digital certificates. The CP typically addresses: the community of users involved; appropriate uses of certificates; obligations of all parties; publication of certificates; privacy and confidentiality provisions; general requirements for CA operations including application procedures; identification and authentication; certificate issuance; certificate acceptance; renewal, rekey, and revocation of certificates; and other legal provisions. It also includes a detailed description of digital certificate format at the syntax level. The format is tailored specifically to meet the operational requirements of the participating organizations as described in the Concept of Operations (ConOps).

Certificate Practices Statement

The CPS is a public statement of the practices for issuing and validating certificates and for supporting reliance on certificates. Based upon the guidelines provided by the CP, the CPS is designed to be a simple and straightforward explanation of how the system works and to what standards the operations of the CA may be held by the participants. It includes a detailed description of the identity proofing process, certificate issuance and validity period, and certificate revocation procedures and maintenance. This is the working document that describes every facet of digital certificate policy, generation, use, and maintenance.

1.4 Models of PKIs

The Federal Government, in *Access with Trust*, describes PKI as

> a combination of products, services, facilities, policies, procedures, agreements, and people that provides for and sustains secure interactions on open networks such as the Internet. It is not necessarily a single monolithic entity. It might be a distributed system in which the component elements may include multiple public key infrastructures which are interoperable and interconnected.

The unifying element of a PKI is the certificate policy, which describes the community using the PKI, eligible applications supported by it, and the rules of engagement that all parties must follow. In the context of the statement above, then, there are three primary models of public key infrastructures that are worth reviewing in the context of higher education. They differ with respect to the character of the overall community or communities served, and, consequently, with respect to the nature of the policy authority needed to create and administer the certificate policy. These models are: Closed or Enterprise Model, Network or Community of Interest Model, and Open Model.

Closed Model (Enterprise Model)

Some communities are, by their nature, self-contained with their own well-defined policies. The typical example given is a business in which its employees and management would represent such a self-contained community. In this scenario policies pertaining to identification and authentication of employees, storage of private keys, management of the CA, and so forth can be dictated by the management. In the "closed" scenario, the PKI policies, practices and procedures are of interest only to the local community served. In the closed model, a unit of the enterprise's management organization plays the role of policy authority.

Network Model (Community of Interest Model)

There are also communities that can best be described as a union or conglomeration of multiple distinct communities. Industry associations and trading partnerships are both examples of such communities of communities. For instance, the Automotive Network Exchange (ANX) is a network of independent companies that do business together in the automotive market. Another example might be the Digital Library Federation. In the network scenario, a group of organizations come together to collectively identify the policies, procedures, security services and applications that will be supported by their PKI. A PKI deployed to support the interactions between members of such extended communities is defined as a "network" model. The policy authority can be a pre-existing body, as in the case of the ANX, or can be a body created by the participants specifically for the purpose of managing the PKI. The Digital Library Federation is an example of this model and is discussed in the next section.

Plans for network or community-of-interest PKI's in higher education typically center around functional areas like the administration of student financial aid -- a group that represents a subset of all students; a subset of administrators; plus external entities like the Office of Student Financial Aid and financial institutions.

Open Model

Some communities are very broad in nature, and may be associated with policies that are either informal, or largely specified in regulations and legislation. Some examples are the community of all retail consumers, or the community including and supporting the educational needs of all graduating high school seniors. A PKI model serving such broad communities is referred to as "open." Open PKI's are still largely in the conceptual stages, and their difference from a network PKI may be viewed as a matter of degree. Conceptually, however, in an open PKI one entity or group of entities would provide the authority for the issuance of certificates that could be relied upon by a spectrum of different parties. The policy authority could be one organization, championing a particular approach on behalf of a larger community (GSA ACES model) or a collection of organizations developing a shared approach but willing to extend the PKI beyond their initial membership (ABA TrustID example). In these cases, the goal is to create a "portable Internet credential" that has broad utility for identifying oneself electronically -- beyond the originating community.

1.5 A PKI Model for Higher Education

The higher education community is an extraordinarily complex collection which comprises public and private educational enterprises; students; high schools; lending and other financial institutions; federal and state government agencies; employers; and other participants. Each of the models of PKI discussed above has a place in this arena.

The PKI Work Group reviewed a number of critical, frequently recurring and sensitive exchanges of information that occur among students, high schools, postsecondary institutions, guaranty agencies and employers. These involve the exchange of application information, test score data, grades, transcripts, and loan information. If these are to be successfully moved to a public-key based exchange, some type of open PKI will ultimately be necessary, as it will be impossible to fully characterize the participants in real time.

These transactions:

- Cut broadly across constituencies related to higher education, including schools, students, and potential employers
- Are often time-sensitive
- Require authentication and confidentiality

These transactions appear to fall primarily in the arena characterized by an open PKI.

The PESC PKI Work Group recognizes another important ongoing study of PKI issues in higher education – a PKI Work Group of NET@EDU, under the auspices of EDUCAUSE. The initiative supports X.509v3 digital credentials as the de facto standard for institutions. EDUCAUSE, the University Corporation for Advanced Internet Development, the Corporation for Research and Education Networking, and the Coalition for Networked Information endorsed these credentials in a March, 2000, letter to Federal agencies, encouraging use of the standards when dealing with higher education institutions.

1.6 Reference URLs:

For a more complete overview of how digital signatures work and how they are supported by a public key infrastructure, see the following references.

http://www.gits.gov
Access With Trust, the Government Information Technology Services (GITS) Board and OMB's description of how PKI can enable many of the goals laid out in the Access America initiatives.

http://www.rsasecurity.com/rsalabs/faq/
RSA Laboratories' Frequently Asked Questions About Today's Cryptography. This FAQ covers the technical mathematics of cryptography as well as export law and basic fundamentals of information security.

http://www.digsigtrust.com/digital.html
Digitial Signature Trust's website containing useful links to get a demonstration certificate; view a live demonstration of the relationship between a certification authority, subscriber, repository and relying party.

http://www.pca.dfn.de/dfnpca/pki-links.html
The PKI Page, from Germany.

http://www.educause.edu/netatedu/groups/pki/report.pdf
A well-written explanation of PKI technology which appears on the NET@EDU as an output of its PKI Work Group.

2 The Use of PKI in Education and Other Industries

Universities have special authentication and security requirements pertaining to their students that are quite different from a commercial or retail model. These derive from both the nomadic nature of the core community and also from the need for a portable certificate. Students use their student ID's very broadly and for a wide variety of campus-based services (building access, meals, library access), and may use a number of different computers across the campus and at home. Accordingly, the need for a portable certificate, located in a smart card or other token, plays a much greater role for this enterprise than it is likely to in an office- or home-based environment. An important exception is military bases, where there are very similar drivers.

Examples that will be discussed below include University of Texas and University of California.

2.1 The University of Texas System (U.T.) – Closed Model

The University of Texas System provides a good example of an enterprise-oriented or closed PKI in higher education. The University of Texas System Strategic Information Council approved the Phase I Implementation of the University of Texas System Public Key Infrastructure (UTSPKI). Implementation of Phase I began in January, 1999. The University of Texas System consists of 15 component institutions, including four-year undergraduate, graduate, health science centers and medical schools. The UTSPKI is designed to aid the System and the component institutions in accomplishing their missions by making the move into cyberspace much easier.

The University of Texas Health Science Center at Houston will use a "Locally Hosted Option" to issue and manage certificates. This process allows users listed in a University of Texas at Houston LDAP Directory (Lightweight Directory Access Protocal Directory) to apply for certificates from the CA. Once a user's identity has been appropriately verified by a local registration authority (LRA) at the University of Texas at Houston, his or her certificate will be issued.

The remaining 13 components and University of Texas System Administration will use a "Remote Hosting Option" in which users will directly request certificates from the CA. The web interface for requesting a certificate is customized for each University of Texas component CA. Local registration authorities (LRAs) at each of the components and at the System Administration will verify the identity of each individual requesting a certificate. Upon identity verification, the requesting user will be issued a certificate.

On March 25, 1999, a Master Service Agreement was signed with a commercial certificate authority, for a period of three years. Fifty thousand digital certificates were purchased. There was an allocation made to each institution and the University of Texas System Administration. Each institution was notified of their allocation and allowed to request more certificates. A final allocation was made, which completed the distribution of all 50,000 certificates. The University of Texas certificates issued to individuals are "High Assurance Certificates" as recommended by the University of Texas System Information Technology Management Council (4/9/1998) and approved by the Strategic Leadership Council (4/15/1998)

So far, items completed in Phase I include activation of a locally-hosted certificate authority at the University of Texas Houston, activation of a remotely-hosted certificate authority at the University of Texas System Administration, and a UTSPKI training session held on June 3, 1999. Also, a request was submitted on June 16, 1999 to the U. T. System Office for General Council for Opinions relating to UTSPKI usage. On August 17, 1999, a Report of the Digital Signature Task Force – U.T. System Office of General Council – was completed. It can be viewed at www.utsystem.edu/ogc/intellectual property/digsigtf.htm. Examples of online forms that can be signed using digital IDs were place on the web on August 21, 1999, at www.uth.tmc.edu/xorgs/utspki/signing-forms.htm.

The University of Texas has a number of initiatives out of the Office of Information Technology and Distance Education. The overall goal of the University of Texas Information Technology Initiative project is to identify system-wide information technology strategies so that components can more effectively compete in the 21st

century. These strategies provide the opportunity to use information technologies to improve the way University of Texas components achieve their basic missions.

More information on the UTSPKI project may be found at: http://www.uth.tmc.edu/xorgs/utspki/index.htm.

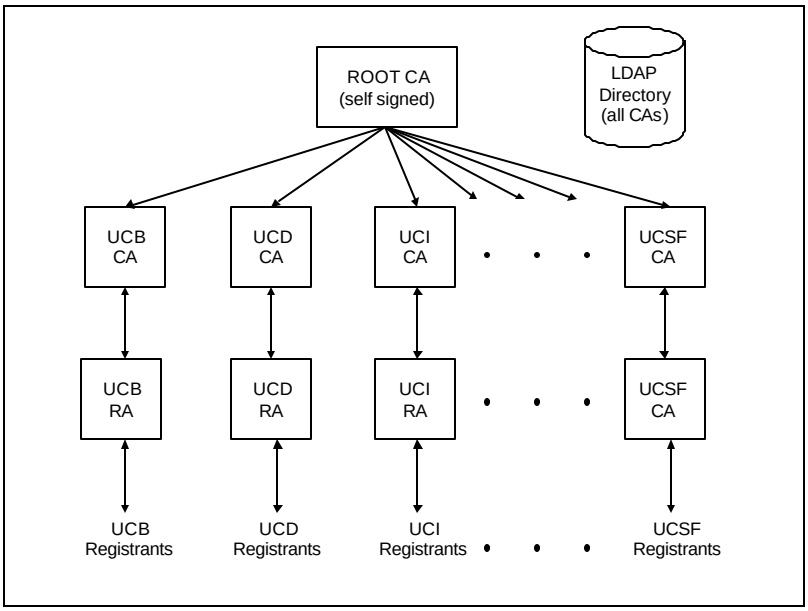
2.2 University of California (UC) – Closed Model

The University of California provides a second example of a closed, enterprise-based PKI. The initial pilot that began in March, 1999 included a locally developed Certificate Authority based on the Netscape CA server. This CA delivered certificates to applicants via a self-service web page interface. Identification was based on existing university systems that utilized userid and password. The CA was in the University of California Office of the President (UCOP) trust chain so that certificates could be used with UCOP applications. Authorization support was not expected to be part of the initial pilot -- applications would unpack the user identifier (UID) from certificates and use existing mechanisms to determine whether or not to grant access. They expected to select no more than 50 testers from a mix of students, faculty and staff. Four initial applications were used during the pilot including employee benefits and an internal staff website.

Authentication was possible in all applications either with certificates or with logonid and password. A general description of the architecture of the initial University Public Key Infrastructure implementation was forwarded to each of the ten campuses of the university. Additionally, each campus was advised of several decisions they needed to make in anticipation of the pilot getting underway this spring.

Initially, all functions will be run on a single platform located at the UCOP Data Center. Separate Certificate Authorities (CAs) will be run for each campus. The Registration function (RAs) will be performed over the network from each campus using standard Web browsers. Each CA will be identified appropriately in any certificates issued so that the CA function can moved to the campus at a later date, if desired.

At startup, the root Certificate Authority will sign a certificate for each of the campus CAs. For the initial rollout period, ten instances of the Certificate Authority (CA) and the Registration Authority (RA) -- one for each campus and UCOP-- will be run on the same server. Registrants or designated Certificate Registrars on each campus will be responsible for interacting with that campus' RA over the network, as shown below:



Two means of initial authentication will be supported, each resulting in a different strength certificate. Using an automated method, authentication will be performed against data in the University Directory. A second,

"manual" process is also supported. Under this method, an individual visits a Certificate Registrar in person, who, after confirming the physical credentials of the applicant, uses a Web browser to access the RA and obtain a certificate for the applicant. The certificate is placed on a diskette and given to the applicant. The in-person method is the only means of authentication available for use by students.

Certificate information will be published to a lightweight directory access protocol (LDAP) directory devoted solely to this purpose. Again, initially the LDAP directory will be operated at UCOP. Some campuses may want to operate their own local RA or RA and CA. This capability will be supported at some time in the future.

Each location will support at least two roles: a Certificate Policy Administrator, and a Certificate Registrar. The Certificate Policy Administrator is responsible for specifying local certificate and enrollment page contents, managing reporting, and any location-specific configurations. Following the chain of certificates hierarchy, the person in the Certificate Policy Administrator role receives his/her authentication from the root Administrator at UCOP.

The Certificate Registrar performs local certificate management functions. These include manual approval and rejection, face-to-face (manual) authentication, revocation, and generally managing day-to-day certificate operations. In other installations, the Certificate Registrar's role typically is performed by someone from the HR or Student Registrar areas (or one from each). The Certificate Registrar's access privileges are granted by their local Certificate Policy Administrator and access at the workstation level is authenticated via smart card. The only client software necessary is a browser. There is nothing to preclude a campus having one person play both the Certificate Registrar and Certificate Policy Administrator roles, and some may find this advisable during the pilot. For more information, refer to the following URL: http://www.ais.ucla.edu/auth/index.htm.

2.3 Automotive Network eXchange (ANX) – Network Model

The Automotive Network eXchange (ANX®) is a secure and reliable global network infrastructure that replaces complex legacy networks and facilitates the re-engineering of supply chains within the automotive industry. The ANX, introduced by the Automotive Industry Action Group (AIAG) in 1995, was designed to improve on the inefficiencies, network inconsistencies and lack-of-data-security offered by the public internet. The ANX securely connects trading partners electronically – allowing them to collaborate on product design and development; solicit and process orders; facilitate just-in-time manufacturing, coordinate purchasing, and post shipping schedules. ANX provides a safe, robust and neutral trading environment for companies to communicate, collaborate and facilitate supplier relationships.

In March of 1998, the AIAG selected the first provider of certification authority services for the network, which is designed to meet the highest degree of security and privacy for data through using digital certificates.

ANX Service is separate and distinct from the public internet. ANX Service is delivered over a virtual private network established solely for the transmission of ANX Traffic. This virtual private network (VPN) is not shared with any other users. ANX Service uses the TCP/IP protocol for networking but uses facilities that at a minimum are logically separate from the facilities used by the public internet. The Trading Partner's ANX Interface may carry traffic destined for both the public internet and the ANX Network.
The first internet protocol security (IPSec) certificate was issued in the ANX production environment in September 1998 by Digital Signature Trust (DST), ANX's Root CA. Through its TrustSource Plus Certification Authority Service, DST issues certificates to ANX trading partners and manages all aspects of the certificate life-cycle. Parties seeking to rely on certificates tap the TrustExchange repository, the DST-managed database for ANX certificates, to check a certificate's status. The repository enables trading partners to communicate over the ANX Network and feel assured about the identity of the sender and the privacy and integrity of their transactions.

2.4 The Digital Library Federation – Network Model

The Digital Library Federation (DLF) was founded in 1995 to establish the conditions for creating, maintaining, expanding, and preserving a distributed collection of digital materials accessible to scholars, students, and a wider public. The Federation is a leadership organization operating under the umbrella of the Council on Library and Information Resources. It is composed of participants who manage and operate digital libraries. The DLF's efforts in using PKI presume the need to rely upon certificates from multiple independent CA's, so it is a more complex example of a networked approach. Broad success will hinge heavily upon surmounting both policy and technical interoperability barriers.

Under DLF auspices, the California Digital Library, Columbia University, JSTOR (Journal Storage, an on-line repository of scholarly journals), and OCLC (Online Computer Library Center) have developed a protocol that enables an information resource provider to verify that a user bearing a digital certificate has authority from a home institution to use a requested resource. The prototype system developed combines the use of X.509 digital certificates for authentication with a directory service providing authorization to licensed resources based on user attributes. An architecture statement has been developed as a result of the prototype development work as well as a project overview. It should be noted that at least one of the organizations mentioned above utilizes IP authentication for web access; this is a clear instance where use of digital certificates would improve service to users.

The group has also made the assumption that each institution will have its own certificate authority (CA). Thus, the information contained within the certificate is sufficient to identify the institution. Clearly, the architecture will need to address the case in which the institution is not also the CA, possibly by requiring that the institution be identified in a designated field within the certificate. Another assumption is that the full authentication and authorization process is performed infrequently (e.g., once per "session") so that directory load can be minimized.

One of the key design decisions made early in the design process was the separation of authentication and authorization requirements. Authentication for the transmissions, messages, and their originators required a different processing than that for authorization, which was to ensure that the user was authorized for access to the system, application, and its underlying data. The criteria of localization of information, accommodation of temporal change, and privacy considerations led to the conclusion that authorization information cannot be explicitly included in the certificate payload. Thus, the institution must have a directory or attribute server which, given some information from the certificate, can determine eligibility for the service. To simplify the directory and access protocol discovery, it was decided to place a URL encoding the query in one of the X.509 certificate attribute fields. The service provider does not need to interpret the contents of this URL beyond interpreting it as a URL.

For more information, please refer to the following URL: http://www.clir.org/diglib/

2.5 GSA ACES – Open Model

The U.S. General Services Administration (GSA) is sponsoring an initiative called Access Certificates for Electronic Services (ACES). Much of the impetus behind ACES flows from the Access America initiative championed by Vice President Al Gore. The primary goal of Access America is to leverage information technology to deliver comprehensive government services to Americans and to dramatically increase government productivity. In line with this goal, ACES is meant to facilitate the public's access to government services via the Internet.

ACES is intended to enable individuals and businesses to efficiently conduct business with the government electronically via the use of digital signature technology. Government agencies, individuals, and individuals acting on behalf of businesses may receive digital certificates that positively identify them online. ACES certificates will enable agencies to authenticate users prior to granting access to confidential, personalized information. Users gain assurance that their identity will be validated before personal information is released, and that the information has not been tampered with in transit.

As a result, government agencies will be able to provide more personalized services online, accept various kinds of document filings online, and make accessing federal government services faster and more convenient. A citizen who gets an ACES certificate can use that same certificate to interact with any federal agency participating in the ACES program, an example of the “open model” PKI. In this case GSA has provided a policy authority on behalf of the Federal government. GSA created a certificate policy designed to meet the needs of the largest possible number of Federal agencies needing authenticated communications with their constituents. In addition, GSA developed, or sponsored the development of, the needed supporting contractual documents, including contracts with each of the three ACES vendors; the vendors' certification practice statements, which support the CP; qualified relying party agreements entered into by participating agencies, and subscriber agreements for participating citizens and businesses wishing to use an ACES certificate. This constitutes a broadly open Federal-purpose PKI that agencies may opt into as they become prepared and have a need.

2.6 ABA TrustID® Program – Open Model

The banking industry has also stepped up to sponsor the creation of a trusted online identity credential based on public key cryptography, analogous to an ACES certificate but with a broader goal. Where ACES is intended for Federal purposes only, the bank-sponsored credential, called a TrustID certificate, is designed to be broadly acceptable across the Internet economy. Banks will issue TrustID digital certificates to business and consumer customers to support the projected large volume of business-to-business (B2B) and business-to-consumer (B2C) e-commerce.

The American Bankers Association (ABA) is sponsoring this online trust initiative and is hosting a central repository for all bank-issued credentials that is backed by a set of governing policies it manages. In this way, the ABA is adapting its traditional role of providing trust to the needs of businesses and individuals on the Internet.

In addition to its more traditional role, the ABA is dedicated to facilitating electronic commerce by creating a secure environment for financial institutions and their customers. The TrustID certificates that ABA-member banks will issue to customers are designed to bring "portable trust" to Internet e-commerce in the same way that the addition of the Visa brand changed a local bankcard into a trusted, international payment tool. Where the Visa brand promised a "guarantee of Payment," the TrustID branded certificate promises a "guarantee of identity."

This initiative represents the broadest example of an open PKI in that TrustID certificates are explicitly intended for broad use across the economy. Organizations wishing to rely on TrustID certificates do not have to be part of the existing banking community or be one of the founding organizations, but have only to enter into a relying party agreement under the TrustID Certificate Policy.

For more information, refer to the following URL: http://www.digsigtrust.com/fs.html.

3 Issues in Deploying and Using Public Key Technology in Higher Education

3.1 Acceptance by Consumers

The public at large, and students in particular, have shown great willingness to use the Internet and other electronic means to transact business. Hesitancy to shift from traditional to electronic services typically centers around four arenas: access to the technology; ease of use; access to technical support; and, increasingly, trust. Since trust also addresses areas of privacy and confidentiality, it will be further discussed in section 3.2.

Access to the technology

Access to the Internet is probably not an issue for most schools, postsecondary institutions, employers, and other institutional participants, since most of the services of a PKI are or will become available from any current browser. While probably anyone who wants it can get access to the Internet through their school or a public library, delivering and using certificates on a publicly shared computer still presents a lot of customer service support challenges. The customer service aspect may be mitigated by the use of smart cards, but smart cards require a substantial investment in hardware and software.

Ease of use

Most of the desired security services are available using standard Microsoft or Netscape browsers and email clients. Those services not readily available can typically be achieved using plug-ins, and there is an expectation that as markets develop for these, the major browsers will incorporate them. However, requesting, downloading, protecting and managing a certificate will be fundamentally new tasks for most people, and many may hesitate to commit to converting a business transaction from something known to something completely unfamiliar. Adoption rates may be expected to be fastest among communities that practice some form of electronic data interchange (EDI), have already been exposed to certificates in another context, or among students who are less committed to how things have been done in the past.

Technical Support

In light of the ease of use issues, technical support will be critical to the success of any PKI. Further, since Microsoft and Netscape manage certificates very differently (and Microsoft IE handles certs differently from one browser version to the next), tech support offerings will have to be quite sophisticated.

3.2 Trust

Notwithstanding the public's willingness to adopt electronic business practices, there is an expectation that electronic transactions should have, and should be perceived to have, a measure of security that is as good or better than that provided by the prior paper-based system. In the white paper, “Why is Certification Harder Than It Looks?,” by Ed Gerck (http://www.mcg.org.br/whycert.html), Mr. Gerck attests that a secure design is not the same as having trust in the protocol, trust in the correctness of the PKI model, and trust in the effectiveness of PKI. Two basic areas of trust, privacy and confidentiality, are very large concerns for the general public

Confidentiality

Confidentiality is the protection of transmitted data from passive attacks, such as unauthorized monitoring of the application or its data or transference of the data over the network, and the protection of network traffic flow from analysis. This requires that the information in the computer system and information that is transmitted is accessible only by authorized parties and the information is not disclosed or revealed to unauthorized persons. Therefore, it is incumbent on the CA and the application providers to assure that the applications are safe, that the use of certificates will provide strong authentication of the participants, that transactions cannot be intercepted, altered, or viewed in transit, that the data collected is held in a safe

environment, and that there are sufficient means available to the certificate subscribers to protect their own keys.

Privacy

Privacy is a concern that has to do with the use of the information gathered in support of the issuance of certificates, and in the applications themselves. This nation has a deep, long-standing commitment to the concept of personal privacy, and so anything that looks like a "national identifier" is suspect. So, too, is the notion that information collected for identification and authentication purposes might be sold on the commercial market. It will be incumbent upon CAs, and similarly upon education and governmental applications that collect personal information, to communicate very clearly that they are in the *trust* business and not in the *information* business.

Business communications which are sent over traditional means, US Mail or proprietary networks, are protected by laws as they travel from sender to receiver. However, business communications that travel over public networks such as, the Internet, are not protected by federal or state law. Organizations that use the public network to transmit business communications must ensure they take the necessary steps to satisfy their customers concerns for data integrity and confidentiality.

3.3 Interoperability

Interoperability is a concern when using any technology including PKI. The major areas of concern for interoperability are with the various standards, the actual technology, and the policies created to support the PKI. Also, note that initiatives and pilots are currently addressing these challenges. Each of these is discussed in the following paragraphs.

Standards

There are many groups involved with standards creation and promulgation for PKI. Some of the groups that are working on formal standards in these areas are the International Standards Organization (ISO) X.500, the Internet Engineering Task Force (IETF) working groups, and the American National Standards Institute (ANSI) X9. A set of standards put forth by RSA Laboratories dubbed the Public Key Cryptography Standards (PKCS) serve as de facto standards within the cryptographic and PKI communities. The National Institute of Standards and Technology (NIST) also maintains a number of cryptographic standards, called Federal Information Processing Standards (FIPS) and coordinates validation programs for many of these standards. Other relevant efforts include the Federal PKI (FPKI) Steering Committee, the Department of Defense PKI. Standards have also begun to mature for three of the application areas with the broadest initial impact: secure email, secure web access and secure remote dial-in. However it is important to recognize that the existence of a standard, and even compliance with those standards, is a necessary but not sufficient condition for interoperability.

Technical interoperability

Despite the fact that we have a standard in the form of X.509 v3 certificates, certificates issued by one CA are not routinely recognized by another. The standards offer a great deal of flexibility in how they are implemented. Interoperability is achievable but requires commitment and investment on the part of all participants.

Policy interoperability

A critical premise of the NACHA study below is the primacy of policy considerations over technical ones. The NACHA team addressed technical and policy issues simultaneously, but concluded that the creation of a policy authority should have been the first step, and that a policy authority and its products (certificate

policy) are foundational to the ability to interoperate at a later time. This is the approach that ANX took in establishing its PKI.

Other Interoperability initiatives: Bridge CA and NACHA

Two other PKI initiatives and pilots have also been underway to support interoperability between technology and vendors.

Federal Bridge CA

The Federal PKI Policy Authority facilitates interoperability through the Federal Bridge Certificate Authority (FBCA). The FBCA focus is to determine the certificate policy mappings between various governmental CAs. The project is established such that all agencies that interoperate through FBCA are voting members for the project. Interoperability of the CA's through the FBCA is not a requirement, but it an attractive goal.

In order for this project to succeed some boundaries had to be established. The first set of boundary conditions is with respect to the technology. The second boundary conditions were policy and political boundaries.

The technology boundaries are based upon standards. The first of these is compliance with FIPS 140-1, 186. This required a level 3 crypto module for FBCA. Another standard base was to meet the MISPC to maximum extent practical. The final standard base was compliance with X.509v3 certificates. This included the mapping the certificate policy of these certificates as a method of ensuring the success of FBCA. One last technological requirement was the use of commercially available software and hardware.

The political and policy boundaries were based upon a desire for a solution, which was as fully "inclusive" for vendors as possible, meaning that this was not a "single or sole" source CA solution. The FBCA has to support four levels of assurance: rudimentary, basic, medium, and high. The certificate policy generated for the mappings is analogous to the certificate policy (CP) used in Canada. Again, it is important to note that interoperation between the various CA's is not a mandatory goal or requirement of the FBCA .

National Automated Clearing House Association – NACHA

According to the results of the Certification Authority Interoperability Pilot conducted by the Internet Council of NACHA - The Electronic Payments Association, banks can serve a value-added role in electronic commerce by acting as certification authorities on behalf of their customers. The results of the five-month pilot are documented in a NACHA publication, *Certification Authority Interoperability: From Concept to Reality*, available at http://www.nacha.org.

3.4 Key Storage

Key storage is a particular problem in the higher education environment which includes students who are mobile and may use numerous computers for various transactions. Machine-specific solutions for key storage are not practicable in such an environment. It has been suggested that directory storage of private keys is an option, although trust in the security of such a directory is required.

The appearance of smart card readers on major brands of laptop and desktop PC heralds a new era for authentication at the desktop, suggesting that smart cards may be the answer for a mobile population. Several major PC manufactures are beginning to incorporate this technology into their products.

The smart card industry still faces issues of standards compliance, and other thorny issues relating to PKI. For instance, keys generated on a crypto smart card cannot be backed up by the PKI -- the card simply will not allow the keys to be exported since it would violate the premise of non-repudiation. However, it introduces an important single point of failure -- if users lose their smart cards they lose their private keys and their digital identity.

3.5 Certification Authority Services

The UT Austin and University of California examples above illustrate two models for obtaining certificate authority services. As an institution determines its role in a developing PKI and what services should be outsourced, certain areas deserve careful scrutiny. Some areas to be addressed are the knowledge of the technology and its implementations, the policies and procedures under which the certificate authority will operate, the data processing requirements of the certificate authority, including the hours of operation, repository needs, and key recovery implications as well as the facility of the certificate authority, the amount of risk and liability to be borne by the certificate authority, the applications to be supported by the PKI and the underlying use of the certificates, and customer support concerns. Each of these areas are discussed in the following sections.

Technology Knowledge and Implementations

Underlying knowledge of the technology and its implementations are paramount to understanding the use of digital certificates. Currently, there is no one standard for what a certificate means or for what is to be used or for how it is to be used. It is incumbent upon the operators of the PKI and the certificate services to understand how digital certificates will be used within the educational institution and with its partners. Will they be used for authenticating users to an application or to a system? Will they be used to allow servers to authenticate to each other? Will they be used for digitally signing student documents, such as thesis submissions and term papers from students to faculty, or strictly for administrative purposes? Will they be used for encryption? Is there a need to have different levels or classes of certificates? All of these questions and more play an important part in the knowledge of how the technology works and where the technology can be implemented.

Policies and Procedures

Defining policies and procedures are critical steps in using PKI. If digital certificates are used as a means for projecting legal assurances that security risks are controlled, then the scope of these policies and procedures incorporates every security issue. Every common practice, usage, convention, and risk factor must be addressed if these policies and procedures are to be enforced without loopholes. If just one exploitable loophole exists, the integrity of the CA and the infrastructure, which it supports, will be undermined. The vast majority of overt policy and procedure violations typically are borne because of unsecured assets, in this instance the certificate authority, or complacent monitoring and enforcement of these policies and procedures. Violations usually occur from exploiting weaknesses in these policies and procedures. Accordingly, an essential component of policies and procedures is the audit provision that ensures adherence.

Data Processing and Facility Requirements

Within the realm of the CA, the data processing and facility requirements must work in conjunction with other security measures, including, but not limited to, physical security, personnel security, administrative security, and media security.

Physical security is the use of locks, doors, other physical controls, and tamper proofing of sensitive equipment. Measures are taken to secure the facility in which the CA will be housed. Various questions must be asked and answered that will define the physical controls and, thus, the security, of the facility. Access controls protect against unauthorized access to any resource, including the facility. Unauthorized access refers to use, disclosure, modification, destruction, or issuing commands that are not in accordance with identified policies and procedures. In some instances, the use of guards, monitoring devices, locks, and access control devices, such as keys, tokens, and biometric measurement devices are used to secure the facility as well as lead protected walls and ceiling and floor breaks.

Personnel security is the process of screening employees to work in the secure CA facility. Personnel must identify with the sensitivity of the position, be trained to understand and comprehend the security issues related to the task and position, and attend the proper awareness programs. Additional personnel security and controls include the use of background checks, training and retraining, job rotation, sanctions in the event of unauthorized actions, bonding of personnel, monitoring personnel, auditing personnel, and special contractual provisions. The CA has obligations to investigate and oversee personnel who are in trusted positions and to remove them in accordance to the documented policies and procedures if the circumstances warrant. For institutions heavily reliant on student labor pools for information services support, these security targets may be unattainable internally.

Administrative security of the facility includes, but is not limited to, controlling the import of software to the system and within the facility, investigating security breaches to the facility, reviewing audit trails and records, and reviewing the accountability of controls.

Media security for the security includes, but is not limited to,controlling the marking and reproduction of sensitive stored information, ensuring that discarded paper or magnetic media containing sensitive information are destroyed securely, and obtaining proper procedures for secure backup and off-site storage in another secure facility.

Data processing requirements identify how certificates are processed through the CA. These include, but are not limited to, hours of operation, data backups, off-site storage, emergency procedures, and job scheduling. While similar to many other management information systems, these data processing requirements must also include security factors with respect to the physical storage of backup media.

Risk and Liability

The manner in which a CA warranties its services and apportions its liabilities will impact the perception of trust by the CA. In other words, how well a CA guarantees services and determines its liabilities will be perceived by its users that the CA is trustworthy. The CA must be solvent. The CA must have sufficient financial resources to maintain operations and perform its duties in the present as well as in the future. The CA must be reasonably able to bear the risk of liability to users and subscribers. The CA should demonstrate sufficient resources, including insurance, in order to provide reasonable financial responsibility when acting as the trusted third party.

Applications Supported

The complexity of security requirements in modern application protocols have forced a trend toward use of application security measures. In a messaging application, secure messaging demands writer-to-reader protection in an environment in which messages may traverse multiple network connections and may be stored and forwarded through unknown application-level gateway systems. In a web application, two areas of concern exist. The first concern is the compromise of the web server and the second is the compromise of user communications. Both of these web problems need to be addressed through standard application security protocols supported by the web server and browser products that are available from many commercial vendors.

“Home grown” or customized applications that need to have security incorporated into them must be maintained and have a method for recognizing user certificates, validating the user via the certificate and potentially have a need for backward compatibility to recognize expired certificates. Additionally, these applications tend to have a need to be able to communicate with each other – oftentimes between campuses, between client servers, and between clients. It is important to address this during the early stages of PKI planning.

Customer Support

Customer and helpdesk support will need to be available, especially for a new system. Hours when customer support is available must be determined based upon when the application is used. Training on the application must be provided to the people who are answering calls to the helpdesk. An application, such as Remedy, should be used to track the number of calls, the type of calls, and how the call was resolved. People who work on the helpdesk must be appropriately screened for tact and diplomacy, especially when dealing with frustrated users.

User documentation for the application will need to be generated and distributed. Ideally, the design of the application will allow that its use will be intuitive to the user. However, in some cases, written documentation for using the CA and its supporting applications will need to be written, tested, published, and maintained over the lifetime of the CA and supporting applications.

Cost of PKI

The table below compares costs of an organization providing its own certificate services and costs of outsourcing certificate services.

	Infrastructure	Liability	Policy Development	Directories	Updating Applications
In-House	High	High	Low	High	Medium
Outsource	Low	Low	Medium	Low	Medium

Previously mentioned considerations of technology and its implementations, policies and practices, data processing, risk and liability, application support and customer support can be very cost prohibitive for the normal institution. Not only must these costs be borne initially but they are recurring in terms of recurring personnel training, enforcement of the policies and procedures, and helpdesk support. Additionally, many institutions are not equipped to handle the potential damage to their reputations if a lapse in security occurs, private keys are made available publicly, or other infraction of the policies and procedures occur.

While the preceding paragraphs address the implementation concerns of managing certification services in-house, the primary questions an organization must answer and address are:

1. How much verification is necessary? Is the real need to identify a person or organization, or a relationship between the educational institution and the certificate holder?
2. Does the educational institution need control over the issuance and the revocation of the certificate?
3. Does the educational institution want to manage the process and the implementation internally?

With the emergence and widespread use of new technologies, such as PKI, it is important that individuals be able to use a given digital signature within multiple settings and avoid the need for different digital signatures for different groups. Higher education institutions, for example, might want digital signatures that can be recognized and used within the institution's various offices; with state higher education offices; with other institutions; with a variety of Federal agencies providing student, research, and institutional support; with professional associations; and with business, legal, and financial institutions.

Today, there is neither one technology nor one standard. Various organizations, such as the Federal government, are working to achieve these goals for their own business needs, as exemplified by the General Services Administration's ACES contract.

Thus, it is important to promote standards and interoperability and consider carefully before embarking on an approach that may not be sufficiently flexible to achieve these goals. Higher education institutions can either create their own certificate services or use commercial services. To the extent that the latter may already have addressed standards and interoperability – and considered the implications of evolving technology – it may be advisable for institutions to reach the goals of interoperability by outsourcing certificate services.

4 Workgroup Recommendations

The PKI Work Group recommends that PESC formally support and promote the use of PKI for promoting secure electronic commerce in higher education. This would include:

- Posting a statement of support for PKI on the PESC web site and a discussion of the importance of standards in the use of PKI;
- Dedicating an area of the PESC web site for information and materials on PKI and its use in higher education;
- Tracking digital signature legislation in Congress and at the state level;
- Providing support for and cooperating with initiatives to employ PKI technology in higher education.
- Recommending transactions that would be ideal candidates for PKI.

There are a number of organizations within the postsecondary community that have an interest in working toward the consistent use of PKI in higher education. The focus of these organizations varies. Some, such as the Corporation for Research and Educational Networking (CREN), are interested in providing interoperability among university enterprises by providing a common root. Others, focused on subsets of the university business enterprise and their business partners, are centered around functional areas such as financial aid or research grant management.

The Work Group, recognizing the unique position of PESC with its enterprise-wide membership, suggests that its primary role in this issue should be to promote appropriate uses of PKI. To do so, PESC should serve as a monitor of PKI activities in higher education and a conduit for sharing information on these activities. An example might be to follow the institutional and Federal acceptance of X.509v3 digital credentials and make that level of acceptance known to the higher education community. In this way, PESC might facilitate ideas that lead to demonstrations that could develop into larger efforts to create a PKI. PESC might thus bring together various parties who could work together on PKI applications in areas of mutual interest.

Through its study of PKI for higher education and its potential for use in the higher education community, the Work Group recognized that this paper naturally raises a number of unanswered questions. For example, what opportunities exist for PESC investment and/or involvement in a network or open PKI for higher education and what are the likely returns for the membership? What assumptions and hypotheses are associated with PKI that need to be validated, tested, or researched? What are the gaps in the infrastructure that PESC members should be aware of or actively engaged in resolving? What gaps in our knowledge need to be researched and articulated? Are there specific practices that PESC should recommend to its members?

To accomplish the needed promotional, tracking, technical and information-sharing activities, the Work Group proposes that PESC establish a standing committee and contract with a PKI expert to continually survey PKI activities in the higher education community and make that information available to members through published articles, web links, and other electronic communication. With this support, the Work Group believes that PESC can truly further the use of PKI in higher education.

Student Identifier Work Group

Report of the Student Identifier Work Group

A Publication of the
Postsecondary Electronic Standards Council
Washington, DC
May, 2000

Table of Contents

Student Identifier Work Group Members

Dave Stones, AACRAO, Chair
Bill Adams, Sallie Mae
Ellen Blackmun, NASFAA
John Gould, USA Group
Elizabeth Hodgkin, UNIPAC
Laverne Knodle, NACUBO
Carole Kuriatnikova, US Dept of Education
Roxie LaFever, NCS
Margaret Parmalee, USA Group

~

The Postsecondary Electronic Standards Council
One Dupont Circle, NW, Suite 520
Washington, DC 20036
(202) 293-7383
http://www.StandardsCouncil.org

© May 2000

Memorandum

May 5, 2000

TO: PESC Membership

FR: PESC Steering Committee

RE: Follow-on of the Report of the Student Identifier Work Group

We gratefully acknowledge the study undertaken by the Student Identifier Work Group, fulfilling its charge to review student identifiers currently in use and assess the advantages and disadvantages of each for consideration as a single identifier across the higher education enterprise. We believe this study is a foundation for future discussions of the needs and benefits of a single student identifier.

It is our recommendation to the membership that we now set in place a deliberative body to consider the following issues:

1. What are the current and potential future business needs for the use of a single student identifier? What market forces are at work (for example, more movement of students between institutions, more reliance on distance education) that might make a single student identifier more attractive or necessary as we go forward? Whose interests would such an identifier serve?

2. What organizations are likely to take on a single student identifier initiative and how should PESC participate in such an effort?

3. What options are or may be available in the future to meet the identified business needs? Assess the pros, cons, potential benefits and estimated costs of each option.

4. Should there be a separate education identifier or one that is also used for other purposes outside of education? In either case, where and when does the student get his or her first identifier assigned?

5. How can we link the concept of a student identifier to something other than a number, i.e., a thumbprint, a digital signature, a voice print, etc.? What other technology is available to serve the purposes identified in #1 above? Does the selection of technology allow for a replacement identifier and/or the combining of multiple identifiers across participants in the event that an identifier becomes lost or multiple identifiers are given out in error?

We propose the formation of a new PESC work group to respond to these and other questions as a way of informing the membership and setting in place a process for considering whether the Standards Council wants to pursue a solution to the perceived need (if any) for a single student identifier.

I. Introduction and Background

The Student Identifier Work Group (the group) charge was approved by the Postsecondary Electronic Standards Council (PESC) in September, 1998. Reasons for PESC interest in a single student identifier were:

- New technological systems need keys, unique identifiers, to enable accurate matching between persons and documents. The group did not want to encourage a plethora of new student numbers created as new agencies hired new consulting groups to implement new initiatives, as there are currently dozens of different numbers in use, none of which is useful for authentication or for matching with the possible exception of the SSN.
- The Social Security Number (SSN) was sensitive and usage was legislatively restricted

The group believed it necessary to strike before more silos were created and to clarify the issues of greatest importance to universities and their data trading partners.

The group first met in October 1998. It was asked to consider student identifiers currently in use within higher education by institutions, participants in student aid delivery systems, testing services, and service providers. We were to consider identifiers which might have potential as a single student identifier to be used when appropriate across the higher education enterprise. Advantages and disadvantages of various identifying numbers were to be investigated and documented. The group was charged to produce a report informing policy makers of the pros and cons of use of potential single identifiers.

II. Properties and Factors of Interest

As shown below, the group developed a list of identifier properties and factors of interest that would help in the analysis of each identifier under consideration. This list could then aid in the comparison of identifiers and their suitability for serving as a single student identifier.

- **Issuing Agency**—What agency or organization is responsible for issuing an ID to an individual?
- **Format**—What is the format of the identifier: number of digits, alpha only, numeric only, alpha-numeric?
- **Public availability**—Can others obtain this identifier?
- **Sensitivity/Privacy**—Is there some question as to keeping this identifier private? Must it be kept private?
- **Universe**—Define the universe of individuals who may be assigned this ID.

- **Documentation/Certification/Authentication**—Does an individual possess documentation on his/her identifier? Does the identifier certify or authenticate the identity of the individual?
- **Required for internal school use**—Must the ID be used in internal school systems?
- **Required for internal Department of Education use**—Must the ID be used in internal ED systems?
- **Required by other government agencies for education purposes**—Must the ID be used by other government agencies for education purposes?
- **Other uses outside of higher education**—Does the ID have other uses outside of higher education?
- **When assigned**—What event triggers the assignment of the ID?
- **Where assigned**—At what location (office or system) is the ID assigned?
- **Permanency**—Is this ID assigned permanently or can it be changed?
- **Primary intended purpose**—For what purpose was this identifier originally created?

III. Specific Identifiers

The group studied four specific identifiers: Social Security Number, a pair of internal university identifiers, and an identifier for Access America for Students (an initiative to improve federal financial aid processes). The group evaluated them on the same basis, looking at the factors listed in Part II. The Government User ID and Canadian National Student Number were evaluated but not considered as possible U.S. student identifiers. The group determined, however, that logistical issues addressed by these numbers may merit further analysis.

III. A. Social Security Number (SSN)

US citizens and others are familiar with the SSN and its use as an identifier in many arenas. The group found it commonly used in campus student information systems and as an identifier in student financial aid processes. Below is the analysis of the SSN based on the factors outlined in Part II of this paper.

Issuing Agency	Social Security Administration
Format	Nine digit, numeric
Public availability	The SSN is not publicly available, although there are many instances in which the holder is compelled to release it for other uses
Sensitivity/Privacy	The SSN is to an extent a sensitive identifier. There are some sentiments that use of a SSN as a national individual identifier is a violation of privacy, but it can also be argued that the SSN is not "owned" by the individual, and therefore, no personal violation is taking place when it is shared. There are FERPA (Family Educational Rights and Privacy Act) related rules which specifically restrict the use and collection of the SSN, requiring those

	who collect it to make it optional, and to describe how it will be used.
Universe	Available for all born in the US (if they apply), plus all legally admitted aliens
Documentation/Certification/ Authentication	The Social Security Amendments of 1983 (P.L. 98-21) required that new and replacement Social Security cards issued after October 30, 1983, be made of banknote paper and (to the maximum extent practicable) not be subject to counterfeiting. Research is being conducted to determine the feasibility of a permanent type of card material.
Req'd for internal school use	A student must have a SSN to be employed, a recipient of federal financial aid, or reported for tax credit purposes as outlined in the Tax Relief Act of 1997. In many instances a school will use the student's SSN as the key identifier for access to the database file records for the student. When records are matched by other means (name or date of birth), the SSN is often used to validate the match. In other instances, the student may be assigned an internal identification number that may be "linked" to the respective SSN through data processing.
Req'd for internal ED use	The Debt Collection Act (P.L. 97-365) requires that all applicants for loans under any Federal loan program furnish their SSNs to the agency supplying the loan. The SSN has historically been used as a student identifier by ED and loan providers.
Req'd by other gov agencies for education purposes	Other government agencies besides ED use the SSN: for example, the Internal Revenue Service; the Treasury Department; and the Veterans Administration
Other uses outside of higher education	The SSN is used as an identifier in numerous agencies and programs outside of education. For example, the Department of Defense, AFDC, Medicaid/Medicare, state and local unemployment compensation and food stamp programs, Department of Transportation, Department of Housing and Urban Development, Department of Labor, any state blood-donation facility, Civil Service Commission, Office of Child Support Enforcement Parent Locator Service, Selective Service, and all banking, S&L, credit union, and securities institutions.
When assigned	The Social Service Administration (SSA) enables parents to obtain Social Security numbers for their newborn infants automatically when the infant's birth is registered by the State. The program was expanded nationwide in 1989. Currently, all 50 states participate in the program, as well as Washington, DC, and Puerto Rico. All legally admitted aliens are also issued an SSN.

Where assigned	The SSN is assigned by the Social Service Administration.
Permanency	The SSN is a permanently assigned identification number that remains constant throughout an individual's lifetime. In certain circumstances, individuals are allowed to obtain a new SSN if their safety is at risk or if abuse has occurred as a result of another individual's knowledge and misuse of their SSN.
Primary intended purposes	The original intent of the Social Security Board was to accommodate the need to register employers and workers by January 2, 1937, when workers would begin acquiring credits toward old-age insurance benefits. It is still the main intent of the SSA to track retirement benefits, along with disability benefits, Medicare/Medicaid, Supplemental Security Income, and other social insurance programs.

The group identified the aspects of the SSN which support the suitability of the SSN to be used as the single student identifier. Group members felt that the following characteristics of the SSN are advantages:

- Common format.
- Widely recognized.
- Utilized by most governmental and financial entities.
- Available for university applicants before they begin the admission process.
- Assigned to both students and university employees.
- Institutions are required to know and report the SSN, so using it as an ID reduces the total number of student identifier schemes tracked by every institution.

Those identified aspects of the SSN which do not support the SSN as a single student identifier are:

- Eventually, the numbering system will have to be modified to accommodate population (SSA will "run out" of 9-digit numbers).
- Public exposure and privacy issues.
- Foreign students will not always have an SSN.

The group felt that the characteristics and intended usage of the SSN which would make it suitable for the single identifier are:

- Common key identifier for most databases.
- Common individual identifier for other external entities.

III. B. Purdue University ID (PUID)

When information was gathered for this report, Purdue University was preparing to issue a new ID for students--past and present--and current employees. It was anticipated that eventually this identifier would link a student across the internal legacy internal systems at Purdue University.

Issuing Agency	The PUID is a system-generated number.
Format	Nine digits. When placed on an ID card, an ISO number will be used, extending the ID to 16 digits: the first 6 digits represent the entity (Purdue in this case) and the last 10 digits represent the individual in that entity as well as a check digit for verification purposes.
Public availability	The PUID is not available to the public.
Sensitivity/privacy	It has not yet been determined if the PUID should be kept private, although there is a recommendation that it be kept private.
Universe	The universe of individuals receiving a PUID includes any person with some (or potential) relationship to Purdue University. For student populations, this would include prospective students through alumni. The initial implementation is focused on current students and employees.
Documentation/Certification/ Authentication	Since this will be a system-generated number, the student will not be required to show proof that the number is correct.
Req'd for internal school use	The PUID is not currently used in internal systems. It is envisioned that the PUID will be on the Purdue Access Card. As campus systems/services use that card, the use of the PUID will grow. As legacy systems are replaced, it is expected that the PUID will be used as the key, replacing legacy system identifiers.
Req'd for internal ED use	The PUID is not used outside Purdue.
Req'd by other gov agencies for education purposes	The PUID is not used outside Purdue.
Other uses outside of higher education	If third-party vendors wish to offer services to students and base those services on the Purdue Access Card, then other uses outside higher education could be possible. The PUID could be prefaced by an ISO number. (Note: release of the PUID to vendors for that purpose would need to be authorized by the student.)
When assigned	Currently, after a student or employee is added to the legacy system, the information is passed to the PUID system application and is assigned the identifier. Eventually, it is envisioned that the legacy systems will access the PUID system to find a person and obtain the

	identifier. Those systems will then use the PUID as the key supplanting current legacy system identifiers.
Where assigned	The PUID is assigned by the PUID system internal to Purdue University.
Permanency	This identifier is considered permanent. The only exception envisioned will be for an employee who enrolls as a Purdue student and is assigned two numbers.
Primary intended purposes	As the issuing agency, Purdue University will utilize the PUID to uniquely identify people and link multiple identifiers a person may have across Purdue's legacy systems.

The advantages afforded by the use of the PUID are internal to Purdue University. Today many offices keep manual lists cross-listing the identifiers a student has, especially between the student and HR systems to support required business functions. The PUID application will replace those manual lists. It will replace the Social Security Number as the primary identifier. The PUID will provide a standard for departments developing additional database applications.

From a national perspective, the group found the PUID is not unique outside of Purdue University.

The group found the PUID not suitable for a single identifier because of its limited scope.

III. C. University of Texas ID

The University of Texas at Austin (UT Austin) issues an ID to students, faculty, staff and others connected to the University. It is maintained by UT Austin and used by its internal systems.

Issuing Agency	This ID is assigned programmatically by routines on the administrative computing systems at the University of Texas at Austin. These routines are invoked by the UT Austin ID Center.
Format	This ID is a 16-digit ISO number. The leading 6 digits, 600861, comprise a number assigned to the University of Texas at Austin. The last 10 digits allow identification with the University, including a check digit.
Public availability	The ID is not available to the public.
Sensitivity/privacy	At least for students, this number is generated as a result of their admission to UT Austin, and therefore becomes part of their student record. Since it is not listed as directory information, it is protected by FERPA and is private. This number is proprietary to UT Austin and is not as sensitive as other numbers, such as SSN, whose place it

	takes on the UT Austin ID card. This issue was an important factor as UT Austin considered use of this card for financial transactions off campus.
Universe	These numbers are generated for those who study, work, or otherwise operate at UT Austin. This list includes: ▪ Students ▪ Faculty ▪ Staff ▪ Authorized library users ▪ Any other person deemed appropriate to receive an ID card (although generally the reason is to use the library)
Documentation/Certification/ Authentication	The ID number links to other identifiers and subsystems, including the SSN (sometimes authenticated by a personal identifier number (PIN)) and the UTEID (electronic ID), which is used, authenticated by a mixed-format password, for secure web-based transactions.
Req'd for internal school use	Yes.
Req'd for internal ED use	No, this number is unknown to the US Dept. of Education.
Req'd by other gov agencies for education purposes	No.
Other uses outside of higher education	Although this ID is not used outside of UT Austin, there are potential banking-related uses in the future.
When assigned	The number is assigned when the ID card is issued. For students, this is generally during orientation. For staff, it is when they become employed.
Where assigned	The number is actually assigned by a computer routine. Most individuals obtain the UT Austin ID on their ID card, which is issued by the ID Center on campus.
Permanency	The number changes if the ID card is re-issued.
Primary intended purposes	The purpose of the UT Austin ID is to provide an identifier which is unique to the UT Austin community, not as sensitive as the SSN, and which can serve as an identifier meaningful to automated systems and linked internally to various subsystems (Library, Students, Employees).

The Work Group recognizes the advantages of the UT Austin ID:

- Protects against public exposure of the SSN.
- Provides a single number which is available, for and assigned to, every person at UT Austin. Foreign students, for example, are not generally issued SSN's.

However, the disadvantages to this ID are:

- Another number to remember.

- Not used outside UT Austin.
- Not useful for matching records.

Because this ID is so limited in scope, it was found to be unsuitable as a single identifier candidate.

III. D. Access America for Students

A significant amount of work has been spent on an initiative to pilot concepts that would improve the delivery of Federal student financial aid. Called Access America for Students, this effort has now been folded into the full modernization effort of the Department of Education. The group feels that the analysis performed may be useful in future initiatives requiring a student identifier at the national level.

Issuing Agency	The Student Account Manager (SAM) would assign the ID during the initial account setup. SAM will also perform ongoing cross-references with existing student and school identifiers, e.g., SSN, school codes, fund source codes, and DUNS numbers.
Format	The ID would be comprised of 16 numeric digits (private label account number). The first 6 digits represent the Bank Identification Number (BIN #). Then 10 digits would follow which could be assigned based on pre-defined hierarchical definitions (or not) as desired.
Public availability	While Phase I does not require that this account number be provided back to students/parents, the school or fund sources, it could be utilized in future phases. During Phase I, the account number is established when an origination record is received by SAM.
Sensitivity/Privacy	During Phase I and II, the Student Account is for the tracking of data only. The account will not be accessible through normal bank or ATM routes. Privacy will be maintained since additional information and a Personal Identification Number (PIN) would be required for the student to access the account. Schools and the Department of Education will be required to enter a user ID and password to access information at their appropriate security levels.
Universe	The population eligible to receive this ID is limited. During Phase I (Award Year 1999-2000) only 8 schools will be participating in the Student Account Manager pilot. At an average of 10,000 financial aid recipients per school, approximately 80,000 students will have student account numbers. During Phase I, they will not be given the account number even though one is resident within

	the SAM. Up to 50 schools will be asked to participate in Phase II (Award Year 2000-2001). Assuming the same average of 10,000 financial aid recipients per pilot school, 500,000 students will be assigned a student account number.
Documentation/Certification/ Authentication	This is not an issue in Phases I and II. Eventually it is planned that Access America for Students digital signature may be utilized to access account information when readily available and given a significant level of usage (student adoption).
Req'd for internal school use	There is no requirement for school use at this time (except that schools may eventually need to know it for aid recipients).
Req'd for internal ED use	There is currently no requirement for internal ED use of this ID.
Req'd by other gov agencies for education purposes	Access America for Students is a government-wide program sponsored by the following agencies/departments: National Partnership for Reinventing Government, Office of Management and Budget, Education, Labor, Veterans Administration, Internal Revenue Service, AmeriCorps, Postal Service, Park Services, and others. Other agencies are evaluating participation for Phase I.
Other uses outside of higher education	The use of the account number by other agencies or organizations (tied to education) or otherwise is a definite possibility. Such uses would be defined during later phases of the pilot.
When assigned	This ID is assigned during the initial account setup.
Where assigned	It is assigned in the Student Account Manager.
Permanency	Just as with any credit or debit card account, the ID can be changed. For example, if a credit card is lost or stolen, there is a procedure followed to document the loss, review recent transactions, and transfer approved transactions to a new account number (which is linked to the old account number for historical purposes). In addition, for fraud detection, any transactions received on the old account number are monitored.
Primary intended purposes	The primary intended purpose of the ID is to create individual student accounts to allow utilization of proven commercial systems to authorize and track transactions, perform daily settlements across industry participants, and initiate funds transfers across participants and programs.

The group found that as Access America for Students takes advantage of existing commercial systems, the assignment of a student account number allows data from these accounts to be readily accessible for other programs/systems.

Characteristics which make this a possible candidate for the Single Identifier:

- Matching documents. SAM will use the account number in future phases for matching origination and disbursement transactions to account for electronic processing.
- Authenticatable identity. Account number and assigned PIN could be used for authenticating identity.
- The Student Account identifier could be recognized in normal card banking processes, including POS, ATM, EFT, etc. if/when setup within these systems.

However, the initial scope of this identifier is probably not sufficient to allow its consideration as the single student identifier. This number might never extend to foreign students.

III. E. GUIDe Government User ID

Fifteen Federal agencies and departments (for example, NASA, NSF, and NIH) are collaborating to provide a common presentation of the US Government to its grantee organization community. Called the Federal Commons, this effort is designed to support electronic transmission of grant administrative information as part of pre-award and post-award business processes. The Federal Commons supports registration of grantee organizations and users, enabling a single Government User logon, whereby secure access to all participating agencies is provided. The authentication module of the Federal Commons validates the identity of a user accessing the Federal Commons web site. It uses a Government User Identifier (GUIDe) developed for this specific purpose. A user registers his/her own 10-digit alpha-numeric GUIDe through his/her organization, and it is stored in an authentication database. Details of this process may be found at http://www.fedcommons.gov/Docs/.

The documentation available on the GUIDe identifier provides a good overview of logistical support issues. However, the group felt this number would be too limited in scope to use for all students.

III. F. Canadian Government National Student Number

Statistics Canada's new student data collection survey, called Enhanced Student Information System (ESIS--http://www.statcan.ca/english/concepts/ESIS/index.htm), seeks to provide for studies of student mobility, pathways and their relationship to education and labor market outcomes. The ESIS database will contain a single longitudinal record for each postsecondary student in Canada. Maintaining a unique record for each student requires a reliable identification system. For this reason the National Student Number (NSN) has been introduced.

Institutions are encouraged to include a field for the NSN in their administrative systems, as it is to be permanently linked to the student record and attached to the student transcript. The 30-character identifier is comprised of an institution code, a Statistics Canada institution code, and an institution-assigned student number. Initially, this number will be assigned to all postsecondary students at the time of ESIS implementation by each institution. New students to an institution will carry their NSN on their transcript from another institution, or, if they have not yet had one assigned, the institution will assign one.

The group felt it is important to be aware of this effort, study its characteristics, and plan to accommodate this number when it identifies for US institutions students previously enrolled in Canadian schools.

IV. Matrix of Identifiers and Factors

To summarize the findings regarding the first four identifiers covered in Part III, the following matrix was developed by the group. It is hoped that this matrix, along with those issues identified in Part V, will aid comparisons.

	SSN	**PUID**	**UT ID**	**AAFS**
Assigned by:	SSA	Purdue U	UT Austin	Student Acct Mgr
Format	9(9)	9(9) (+6 on card)	9(16)	9(16)
Public	No	No	No	No
Sensitive/Private	Y/Y	Y-*/Y	Y-/Y	Y-/Y
Universe	US-born and legal aliens	Purdue students & employees	UT students and employees	Financial aid applicants
Authentication	Card	None	High security password	N/A
Internal to school	Yes	No	Yes	Yes
By ED	Yes	No	No	Yes?
Other Govt Agency	Yes	No	No	Yes?
Others	Yes	No?	No?	No?
When	Birth or entry	Admission or Employment	Admission or Employment	Aid application
Where	SSA office	Purdue	UT	Student Acct Mgr
Permanency	Permanent	Permanent	Permanent	Changeable

* Y- indicates Yes, but of less concern.

V. Summary of Advantages and Disadvantages

Advantages To Use of Each of Four Identifiers				
	SSN	**PUID**	**UT ID**	**AAFS**
Well known	Y	N	N	N
Widely used	Y	N	N	N
Matching	Y	N	N	N?
Admission applicants	Y	N	N	N
Disadvantages To Use of Each of Four Identifiers				
Format Lacking	Y-	Y-	Y-	N
Privacy	Y	Y-	Y-	Y
Scope insufficient	Y	Y	Y	Y
Serious Candidate for Single Identifier				
Serious	Y	N	N	N

VI. Characteristics Needed, Intended Usage of Single Student Identifier

The group determined that the following characteristics would be beneficial in a single student identifier:

a. Helpful in matching documents.
A student identifier should accommodate the matching of documents for admission applicants who are still in high school and for those who have already attended other universities and received financial aid.

b. Authenticatable identity.
The process of assigning a student identifier should incorporate authentication of the student's identity. The privacy of the student and his/her identifier would be protected since authentication would be required.

c. Label.
A student identifier can be used for purposes of identifying students.

d. Key to a particular system or mega-system.
A student record uses the identifier as the system key

e. Internal linkage to other data bases.
The student identifier can serve as a linkage to other databases within a system.

f. Accessible electronically to users.
Users should be able to access their identifier electronically.

VII. Conclusions

In considering student identifiers, it is important to recognize that universities need internal policies regarding assignment, usage, and protection of Social Security Numbers and other student identifiers. The University of Illinois recently conducted an in-depth study in this area (see http://www.pb.uillinois.edu/Internet/html/ssn_project.html). The resulting policy statement and report, approved in November and December, 1999, contains many points which should be considered by universities and by other agencies when evaluating identifiers.

The Student Identifier Work Group concluded that replacement of the Social Security Number with another number merely adds another number to every system unless institutions are no longer required to report SSNs to others such as the Department of Education. The SSN is often helpful in matching test scores, transcripts, and other student-related documents. An authenticatable ID, SSN or not, makes sense for controlling access and authorization of actions by the student. It may not be useful for matching, and might not be unique unless the SSN is stored behind it.

XML Work Group

Extensible Markup Language For Electronic Transactions in Higher Education

A Publication of the
Postsecondary Electronic Standards Council
Washington, DC
May, 2000

Table of Contents

XML Work Group Members

Tim Pavlick, KPMG, Chair
Ed Elmendorf, AASCU
Graeme Finley, PricewaterhouseCoopers
Ron Hodges, PricewaterhouseCoopers
Rick Jennings, SCT
Darin Katzberg, UNIPAC
Kimberly Koran, CSC
Jesse Mercado, University of Illinois at Chicago
Peg Murphy, ETS
Mary Neary-Morley, PeopleSoft
Pauline Roberts, NACUBO
Jeanenne Rothenberger, AACRAO
Pat Salava, ETS
Dave Stones, Harbinger

This document was produced by members of the XML Work Group in collaboration with technical consultant Michael C. Rawlins, Owner and Principal Consultant with Rawlins EDI Consulting. Mike has over 15 years of experience as a technical consultant in information systems. He is vice chair of ANSI ASC X12 Subcommittee C on Communications and Controls and co-chairs X12C's Future Architecture Task Group which is responsible for technical aspects of X12's work on XML. He is also team leader of the Requirements Project Team of ebXML. Mike is currently pursuing a Masters of Science in Computer Science at the University of Texas at Dallas.

~

The Postsecondary Electronic Standards Council
One Dupont Circle, NW, Suite 520
Washington, DC 20036
(202) 293-7383
http://www.StandardsCouncil.org

©May 2000

1 Introduction

1.1 Purpose & Scope

Over the past year, Extensible Markup Language (XML) has become one of the hottest topics in electronic commerce and the Internet. Some have predicted that XML will rapidly make traditional EDI (electronic data interchange) obsolete, while others downplay XML as just EDI in different clothing. This report has been developed to assist the members of the Postsecondary Electronic Standards Council (PESC), and the broader postsecondary education community, in making sense of XML developments and how they relate to the needs of the PESC community. It presents a future vision for the relationship between traditional EDI and XML, both at the standards development level and the technical implementation level. This vision should assist both standards developers and implementers with developing strategies for using XML.

To achieve this purpose, the report addresses the following topics:

- A brief profile of the postsecondary community outlines the community's needs in relation to electronic transactions
- A brief technical overview of XML and related specifications
- The attraction of using XML for electronic transactions with potential advantages and disadvantages in relation to traditional EDI.
- Representative XML efforts in other industries as well as current XML-based initiatives in education are also discussed
- Developing trends in XML usage and the potential impacts
- Recommendations for PESC action

The report deals with technology issues at a fairly high level, and is intended for both non-technical and technical audiences. Although it is written primarily for the education community, much of the information and analyses is applicable to other industries.

NOTE ON TERMINOLOGY: "EDI" has been traditionally defined as "the application to application exchange of structured business data between enterprises". In this sense, XML may be considered as another technology for EDI, a companion to traditional EDI technologies and syntaxes such as ANSI X12 and EDIFACT. The term "XML/EDI" has frequently been used in this context (notably by the XML/EDI Group). However, many people inextricably associate the term "EDI" with the traditional X12 and EDIFACT approaches. Others think of XML/EDI as basically reformatting X12 or EDIFACT into XML syntax, using the same transaction set structures and data elements. This report intends to address using XML somewhat generically in ways that may have no direct relation to X12 or EDIFACT. So, to clarify terminology, in this paper "EDI" refers to the traditional X12 or EDIFACT approaches. The terms "electronic transactions" or "e-transactions" are used to refer to a larger context which encompasses several different approaches including XML, traditional X12 and EDIFACT, and object-oriented technologies. In general the term "e-business" is used in the same way "e-transactions" is used here. However, in order to avoid any confusion between for profit, not-for-profit, non-profit, and government entities, and any confusion with web-based applications, the terms "electronic transactions" or "e-transactions" are used.

1.2 Profile of the Postsecondary Education Community

PESC supports the postsecondary education community by promoting standards for sharing education-related data electronically. The community it serves is comprised of:

- Public and private postsecondary institutions

- Large and small lenders, guarantors, and servicers of educational loans
- Higher education professional associations
- Software and service providers
- State and Federal agencies

The student data exchanged among these diverse institutions and organizations focuses on student financial aid, enrollment reporting, transcripts and student records, testing results, applications for admission, course catalog information and student prospect information, among other applications. Some of these exchanges have used electronic delivery in closed systems and proprietary standards for a number of years. Others are slowly implementing national ANSI ASC X12 standards over the Internet. Still others, for various reasons, remain tied to their paper-based systems.

Some processes are mandated by a government agency or a single service provider, and those data formats tend to be proprietary. Others have been embraced by an industry group, such as CommonLine, for guaranteeing student loans. X12 standard transaction sets (TSs) have been set and implemented for many processes used by institutions, including:

- Sending and receiving transcripts (TSs 130, 131, 146, and 147)
- Educational Testing and Prospect Request and Report (TS 138)
- Course Catalog Information (TS 188)
- Application for Admission (TS 189)
- Student Enrollment Verification (TS 190)

In addition, supply chain X12 transactions such as invoices and purchase orders are used by a handful of postsecondary institutions.

Other standards of interest to the postsecondary education community include a standard for public key infrastructure, standard data definitions that support electronic exchange of education-related data, standard methods of identifying entities across the education community, and new standards-based data exchange methodologies.

Because of its diversity, generalizations about the whole community are hard to make. Several factors tend to have a great influence on how an organization implements technologies for electronic transactions and other software systems. Reflecting the diversity of the community, these factors tend to be played out differently in different segments of the community.

- Capital Cost Constraints - Higher educational institutions typically have smaller capital budgets for computer software and hardware than similarly sized enterprises in other industries such as manufacturing or retail. This situation is alleviated somewhat by education discounts offered by many vendors, but it is still a major consideration. There is significant use of free, public domain software where applicable. Smaller institutions and lenders may have similar cost constraints, but they may opt for low-cost, off-the-shelf software. At the other end of the spectrum, large lenders such as Citibank, and several government agencies, may have large capital budgets more in line with similarly sized enterprises in the private sector.
- Operational Cost Minimization - Many educational institutions frequently make use of low-cost alternatives to assist in minimizing operational costs. As an example, the higher education community pioneered EDI over the Internet (the SPEEDE/ExPRESS project) to avoid transaction fees charged by Value Added Networks.
- Labor Cost - Some higher education institutions have a fairly low-cost labor supply, often in the form of teaching or research assistants and student projects, to employ on information technology projects. On the other hand, many small lenders and institutions may have little or no programming staff, and little or no budget for outside consulting. Again at the other extreme, large lenders and government agencies may have large staffs with labor costs comparable to private industry.

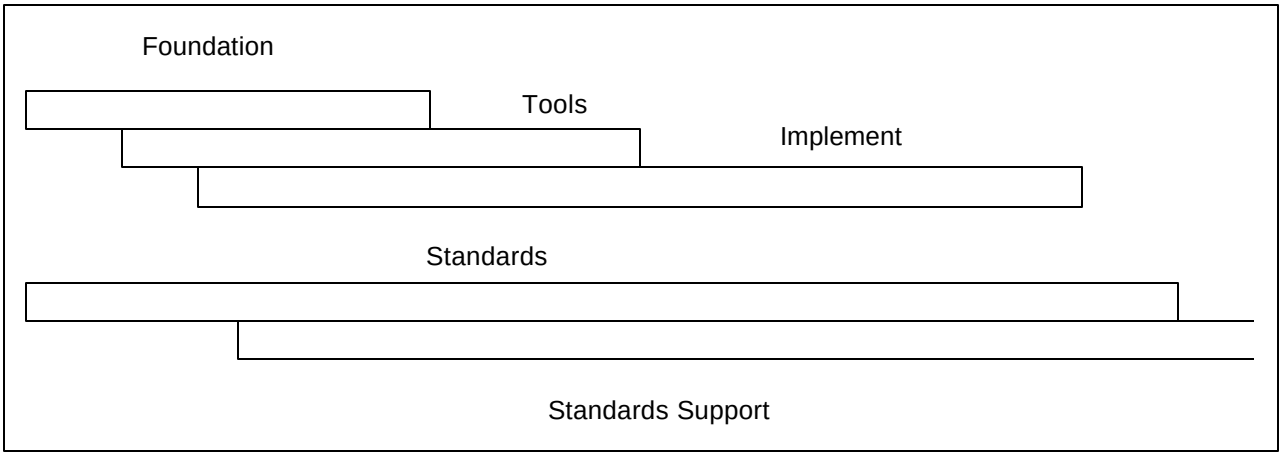
- Limited Market – There are only approximately 3000 higher education institutions in the US in contrast to hundreds of thousands of manufacturing firms and retail enterprises, tens of thousands of which use EDI. This relatively small number tends to limit the market for certain types of application software. On the other hand, certain activities, such as procurement, may be handled by widely available general purpose software.
- Dominance of hubs - In other industries such as manufacturing and retail, trading relationships are typically characterized by "hub and spoke" configurations in which large customers trade with groups of smaller suppliers. The hubs dictate implementation conventions, and the suppliers may trade with several hubs. In higher education, the relationships among institutions exchanging transcripts tends to be more peer to peer. However, some student loan transactions, particularly when government agencies or large lenders are involved, tend to follow the hub and spoke model.

These factors have different influences on the different sectors of the community. For colleges and universities, the capital cost constraints combined with the limited market result in relatively few vendors offering specialized software packages for higher education administration. Due to limited vendor choices, capital constraints, and relatively low-cost labor, many institutions may build their own applications rather than buy them. At present, institutions typically use general-purpose EDI software, but staff rather than vendors often perform integration with administrative applications. However, again noting the diversity of the community, there is a trend toward student information systems vendors embedding EDI translators within their systems. For small colleges and lenders, the same factors may tend to dictate using PC-based commercial, off-the-shelf software, so their preference is to buy rather than build. Large lenders and government agencies may use large commercial packages where they meet their needs but may also perform considerable customization or in-house development when such packages don't completely meet their needs.

The factors noted above will likely hold true and have continuing, similar effects as the technology for conducting transactions electronically evolves beyond traditional EDI.

1.3 The Time Line

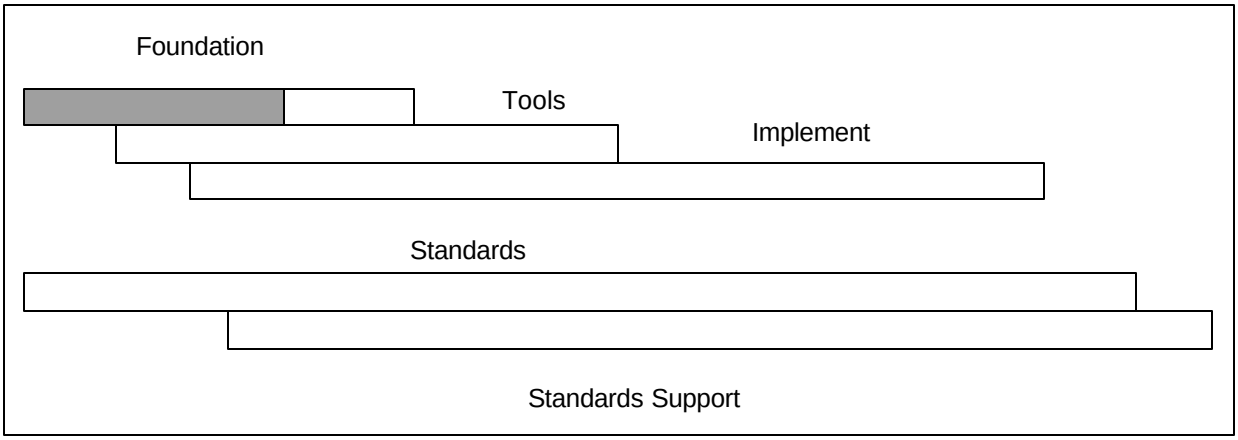
One question revolving around the use of XML for electronic transactions is "When will it be ready?" It is difficult to say for sure, but there are several dependencies that must be met. The time line illustrated below shows the dependencies, and is used as a key to show the range of the time line being addressed and progress along that line in the relevant sections of this document.



1. Foundation - World Wide Web Consortium completion of base XML specifications such as XSL and XML schemas
2. Tools - Tools vendors support for these base XML specifications

3. Implement - Applications vendor implementation of the tools
4. Standards - Development of XML business standards, i.e., Document Type Definitions (DTDs) or schemas for particular business documents
5. Standards Support - Applications vendor support for XML business standards, and widespread availability of easy to use transformation tools

2 XML in a Nutshell



Time Line Progress: Due to publication of XML 1.0 and a great deal of work on related specifications, roughly two-thirds of the foundation work relevant to using XML for electronic transactions is completed.

First and foremost, XML is a markup language. Unlike programming languages, which describe data structures and embody algorithms for processing data, the *primary* purpose of markup languages has been to format and manipulate text. Markup languages were developed for publishing in the days before graphical user interfaces, such as those offered by Microsoft Windows and Macintosh, enabled documents to be formatted with WYSIWIG (what you see is what you get) editors. Markup languages are used to instruct text processing programs that certain "marked up" text is to receive special formatting when printed, such as bolding or underlining. Markup languages, as will be discussed later, have been adapted to other uses beyond publishing and displaying text, but this remains their history and primary motivation.

XML, like it's cousin HTML (HyperText Markup Language), is derived from Standard Generalized Markup Language (SGML). SGML was standardized in the 1980's by the International Standards Organization (ISO). SGML contains a very rich set of features that enable not only a wide variety of formatting but creation and maintenance of things such as tables of contents and indexes. SGML allows the user to create not only single documents but also templates that define the layout and organization of documents. These templates are called document type definitions, or DTDs. Documents can be created compliant with the templates, and when printed all have a consistent appearance and organization. SGML is still widely used in certain industries, particularly in Defense.

HTML was created for the World Wide Web as a single SGML document type. It is then a subset of SGML that is fully compatible with SGML. As web development exploded in the mid 1990's, people began to run into the limitations of this single document type offered by HTML. They wanted to do things with web pages which were becoming awkward with HTML, and wanted additional features such as indexes, tables of contents, and the ability to format a single document both for printing and display using a web browser. However, SGML was too complex for this purpose, and something in between was needed. Thus, the World Wide Web Consortium

(W3C) began developing XML. W3C (www.w3.org) is an international consortium of corporations, government agencies, educational institutions, and other organizations whose purpose is to develop common protocols to promote the evolution of and ensure the interoperability of the World Wide Web. It serves as the de facto standards body for the web, although it produces "technical specifications" instead of "standards" and is not affiliated with or accredited by any other organization such as the ISO or the American National Standards Institute (ANSI). W3C has taken responsibility for maintaining HTML and other web-related specifications in addition to XML.

XML, like HTML, is a subset of SGML. The base XML 1.0 specification was approved by W3C in February of 1998. In computer science terms, XML is not strictly a language but a "meta-language," or a language that is used to define other languages. XML is like a spoken or written language that has an alphabet, punctuation marks, a limited set of grammar rules, and about a half dozen nouns. If it were a language in this sense, you could create words, finish the grammar, and just as easily define the French language as you could English. It is this attribute of XML, its flexibility, that is both its greatest strength and, as will be discussed in using XML for electronic transactions, its most problematic weakness.

The base XML 1.0 specification defines the main features of the XML syntax. Like SGML, it defines both how documents should be created with XML and how to create the templates, or DTDs, that describe a document's organization and the specific elements or "markup tags" used in it. XML looks very similar to HTML, with the tags enclosed in angle brackets ("<" and ">"), and leading and trailing tags surrounding text. Like HTML, two types of information are contained within the tags. One is the XML "element name", and the other is one or more optional "attributes" with associated values. A typical string for a first level chapter heading in bold, blue type, might look something like this in XML:

```
<HEADING level="1", font="bold", color="Blue">Chapter One</HEADING>
```

HTML defines in its one document type all of the elements and attributes that are valid in an HTML web page. XML, on the other hand, defines only a small handful of special elements and attributes. The rest is entirely up to the user. The user may define a DTD for the document that specifies the set of elements, attributes, and their usage that is allowable for the document. However, use of a DTD is optional. If a document complies with a DTD, it is considered "validated" against the DTD. However, without being validated against a DTD, a document may be considered "well-formed" if it follows all of the XML grammar rules.

It is important to understand that XML is not just one specification but a family of related specifications. The XML 1.0 specification only defines well-formed documents and DTDs and does not deal with such things as display and data entry. These are defined in other specifications produced by W3C. Some of the other major specifications in the XML family, and their current status, include:

- XSL, or Extensible Stylesheet Language - This defines a language for creation of "stylesheets" that determine the formatting or display of printed XML documents. This work is still in progress.
- XSLT 1.0, or XSL Transformations - A language for using XSL to define rules for transforming one XML document into another with perhaps different tags, attributes, organization, etc. Intended primarily for use with XSL, rather than general purpose transformations. Approved in November 1999.
- XHTML 1.0, or the Extensible HyperText Markup Language - HTML 4.0 converted to be compliant with XML. Approved in January 2000.
- XML Schemas - More rigorous and detailed means to define XML document templates beyond what is offered in DTDs. This is important for using XML for electronic transactions where business applications may expect data to pass certain types of edits and validity checks before being processed. In progress, but expected to be approved in mid-2000.

Notably missing from the current work list is support for data entry forms, a common feature of many HTML web sites. Some forms support is built into XHTML, and there were early submissions dealing with "architectural" forms and "XML Forms Definition Language."

The divisions noted in the specifications illustrate another of the most important features and strengths of XML in addition to its flexibility. That is the ability, using XML, to separate document content from presentation to or usage by the end user. This enables the same XML document, for example, to be:

- Displayed in a web browser using formatting supplied by an XSL stylesheet
- Nicely printed using an XSL stylesheet
- Directly imported into a spreadsheet or database, perhaps by a future version of Microsoft Excel, for example
- Processed by or imported into a business application, according to the document definition in the DTD or schema.

It is apparent from all of this that XML has great potential for changing not only the web but also many other types of computing applications. However, it is also evident that at present this is just potential. Not only is software not yet available to support the features described above, but the specifications themselves to which the software must conform are still evolving. For example, Microsoft's Internet Explorer Version 5 offers the best XML and XSL support among popular web browsers. However, even it does not yet fully support the latest version of the XSL draft. The authoring tools to create web pages for XSL and XML are still rather limited in availability and primitive compared to HTML tools such as Microsoft Front Page and Netscape Composer. As will be discussed in Section 5, most of the XML software that is currently available is concerned with helping application software vendors build XML capabilities into their products. A reasonable guess is that it may take from one to three years for XML tools to catch up with and surpass HTML tools in both usability and availability. Most knowledgeable people in the industry believe that this will happen eventually, but it certainly has not happened yet. It is also believed that XML and HTML will co-exist for some time to come, even after XML matures.

So, despite all of the hype in the trade press, XML is not yet "ready for prime time." This is nowhere more apparent than in the area of electronic transactions, as a replacement for or next step in the evolution of EDI. Some of the most important base XML technical specifications are not ready yet, and the software that implements them is not mature either. Importantly, despite some claims to the contrary, business standards will be required for XML to reach its full potential for electronic transactions. As will be discussed in the next sections, that work is in its infancy.

3 Why XML for Electronic Transactions?

The previous section gave some hints as to why there is such great interest in using XML for electronic transactions. Indeed, if one keeps up with the trade press, it seems that greatest initial interest in using XML has been in just this area. XML potentially offers some great advantages for electronic transactions over traditional EDI technologies. However, there are a few caveats to be aware of and a few important dependencies to be met before the full potential can be realized. In addition, there are many benefits that have been hyped for XML that may not turn out to be real. Finally, there are some situations in which traditional EDI may still be the best technology for some time to come.

The following figures will be helpful in understanding how XML might be used for electronic transactions and how it differs from traditional EDI.

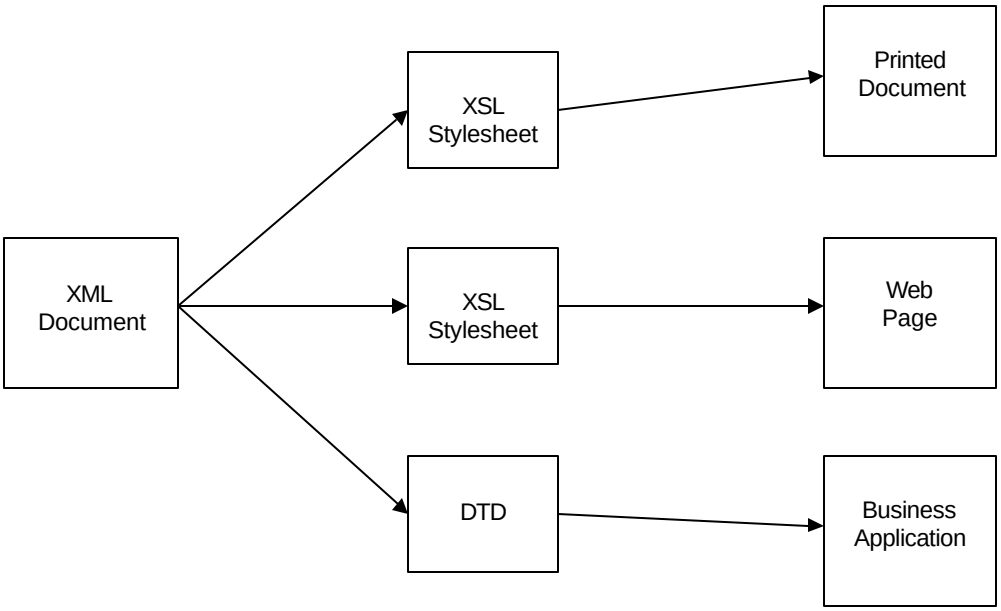


Figure 1 - One Document, Many Uses

As noted in the previous section, one of the goals of designing XML was to separate document content from presentation or usage. This diagram illustrates this, showing that the same XML document may either be printed, displayed in a web browser, or imported into an application. This is one of the greatest potential advantages to using XML for electronic transactions. Like traditional EDI, XML documents may be used by business applications. However, XML differs in its support of small enterprises whose applications may not have an integrated means to process electronic transactions. With traditional EDI, these enterprises must acquire at least desktop EDI systems to print incoming transactions and perform data entry functions for outgoing transactions. With XML, these enterprises may need only a general purpose web browser. Even for those with integrated electronic transaction capabilities, the picture is significantly different, as shown in Figure 2.

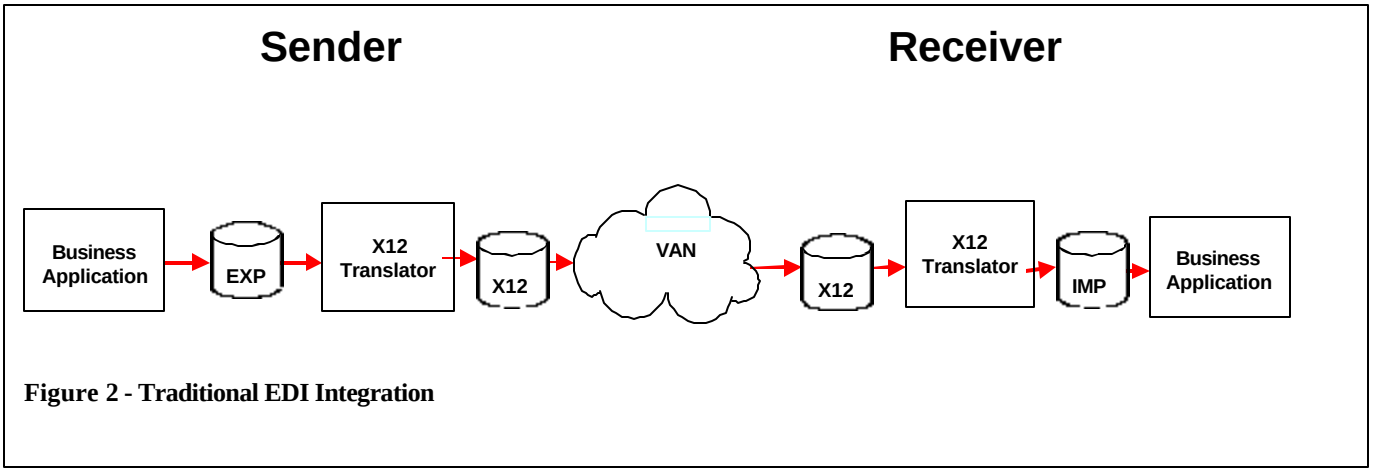


Figure 2 - Traditional EDI Integration

General purpose EDI translation and management software is generally used to integrate traditional EDI with business applications. A processing flow involves the following steps for a typical X12 exchange:

- A business application generates an export file (labeled EXP) with documents to send to one or more partners.
- The export file is processed by the X12 translator to reformat it into one or more X12 interchanges for transmission to trading partners.
- The interchanges are sent through a network to partners. Value Added Networks are still usually employed, although there is increasing use of the public Internet.
- The receiver processes the interchange through an X12 translator to reformat it into an import file (labeled IMP) that may be directly imported by their business application.
- The import file is then processed by the receiver's business application.

XML may make the process somewhat simpler, as shown below.

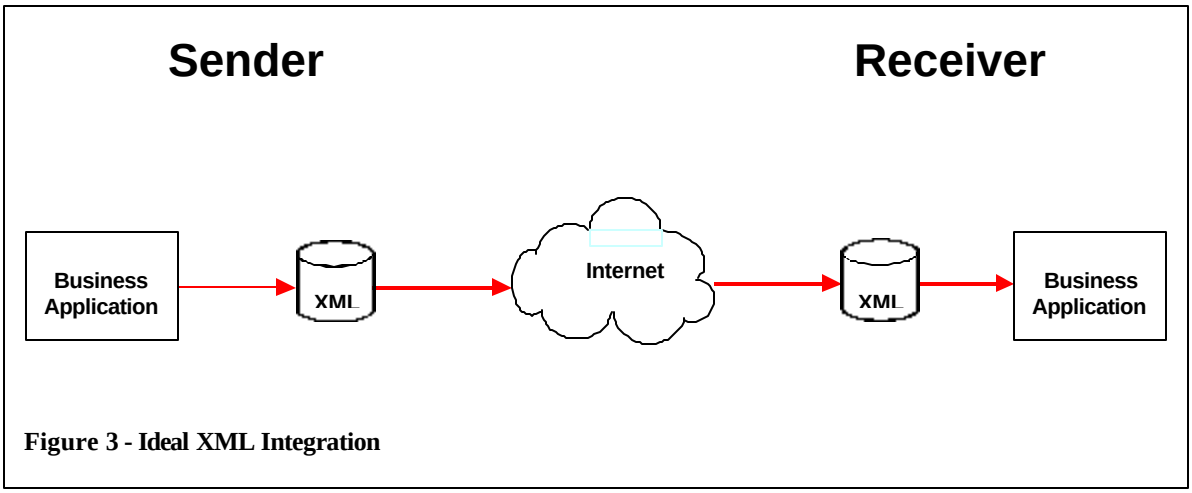


Figure 3 - Ideal XML Integration

The processing flow in this picture is significantly simpler than in Figure 2.

- An XML-enabled business application directly exports a file of business documents in XML format.
- The file is transported over a network, probably the public Internet, to the receiver.
- The receiver's XML-enabled business application directly imports the file and processes it.

At least part of this scenario is likely to occur due to the availability of software development tools that enable XML to be integrated fairly easily into business applications. These tools will be discussed in a later section, but their impact, along with the great interest in XML, is clear. It is likely within a few years that XML will have broad support as a somewhat universal import and export format.

The picture in Figure 3 is a vision that is often hyped by XML evangelists. However, it may be somewhat "ideal" in that it expects each party to agree to the other's format, and that the business applications may generate and process them with no further transformations. This is probably somewhat unrealistic for several reasons:

- Due to lack of standards and the rush to incorporate XML features into applications, the initial implementations of XML import and export features will probably use XML document structures and element names that closely match the vendor's internal structures. For example, a QuickBooks purchase order might use internal QuickBooks field names and look quite different than a PeachTree purchase order. This trend is already becoming apparent.
- Even when file formats are basically the same, due to small differences in the same basic business processes, files exchanged with one trading partner may be different from files exchanged with another. This is quite common in traditional EDI.

It is therefore likely that there will be a transformation at one or even both ends of an exchange, as shown in the next figure.

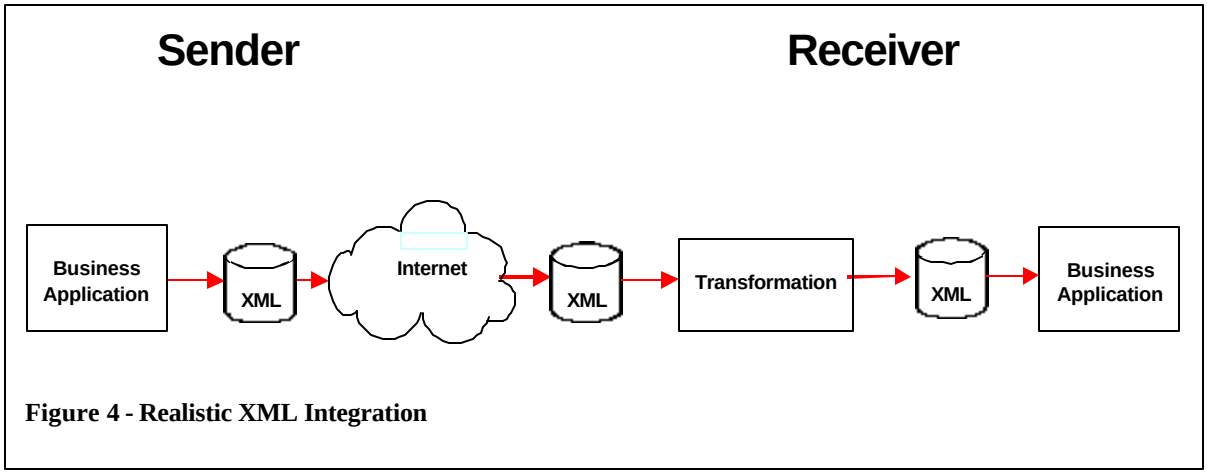


Figure 4 - Realistic XML Integration

This flow is basically the same as in Figure 3, except that there is a transformation at the receiver's end from the sender's format into a format that the receiver's application may process. Conceivably, there may also be a transformation at the sender's end, so that what is exchanged between the two is an intermediate "interchange format" that is directly processed by neither application.

Even though there are transformations in this picture, what is different from traditional EDI is the probability that the transformations may be performed by low-cost, general purpose XML tools rather than specialized EDI translators. The most likely tools to perform these types of transformations are software utilities that use XSLT scripts. XSLT was developed primarily for use with XSL, but there is a good deal of interest in using it as a general purpose document transformation tool. Further, similar to today's graphical EDI mappers, it is likely that there will be fairly easy-to-use tools to create XSLT scripts that perform the transformation from one XML document type to another. Two significant features stand out about these transformation tools:

- They will be general purpose, mass market, and hence low cost. They will support for example, transformations of XML documents for publishing in addition to transformations for electronic transactions.
- The XSLT scripts may be portable, i.e., they may be usable by completely different tools. This is quite in contrast with current EDI mappers and translators where one product's maps can not be used with another. This opens the possibility for users in trading communities to share XSLT scripts. It may also open a market for vendors to provide XSLT scripts with their applications.

An additional advantage of using XML over traditional EDI has to do with ease of use for software developers and implementers. To understand an X12 transaction set, one must refer to the published X12 standards. However, one of the primary strengths of XML is the ability to convey the description of the data along with the data, so that one does not have to refer to a manual to get at least a reasonable understanding of a document. For example, a programmer could read an XML document containing:

```
<STUDENTID type="SSN">123-45-6789</STUDENTID>
```

And know immediately that the field is a student ID number, using the student's Social Security Number. This is a significant advantage, and considering the cost of published standards manuals and the high learning curve in dealing with the standards and implementation guides, is not an advantage that should be discounted.

These are all significant advantages. However, not all of the purported advantages of using XML for electronic transactions are necessarily true. Here are a few that readers would be prudent to examine.

- Business standards for using XML will not be necessary - Strictly speaking this may be true. And, if true, it would be a big advantage over traditional EDI because the processes to develop and approve standards tend to lag far behind the business needs. However, the price of having no standards is potentially a greater number and wider variety of document types depicting basically the same data. XML offers such a great degree of flexibility that the following are only three of the ways that the same data could be formatted:

```
1) <STUDENT_NAME>Fred Flintsone</STUDENTNAME>
2) <STUDENT_NAME>
         <LAST>Flintstone</LAST>
         <FIRST>Fred</FIRST>
   </STUDENT_NAME>
3) <NAME type="STUDENT">
         <SURNAME UN_CODE="BXYZ-23">Flinstone</SURNAME>
         <GIVEN UN_CODE="BXYZ-24">Fred</GIVEN>
   </NAME>
```

  The complete freedom in choosing XML element names, as well as the additional flexibility of using attributes, leads to a near infinite number of ways to format any given item of data. Each different way requires a somewhat different programming approach and transformation. This wide variety of documents means that more transformations must be programmed and supported. This leads to higher labor costs and more error-prone processing. It reduces the likelihood that large groups of users will need similar transformations. This thereby reduces the possibility of shared XSLT scripts or vendors supplying XSLT scripts with applications. In addition, as those familiar with EDI integration will attest, wildly varying transformations may yield some which can not conveniently be done with standard XML tools, and may even yield some which are completely impossible without resorting to other programming tools such as C++ and Java. Standardized element names, attribute names

and values, usage of attributes, and document structures will help to minimize the variations and transformations that will be required.

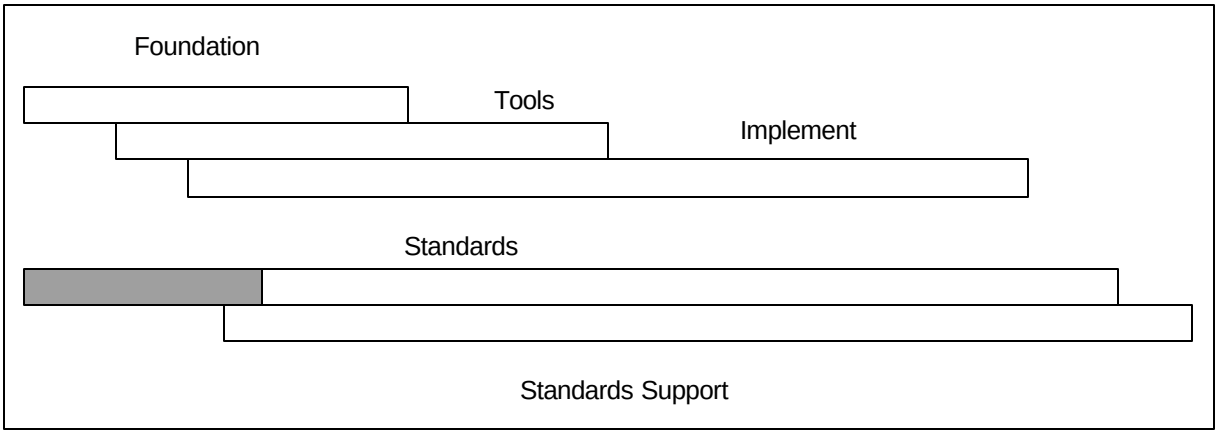
- No translators are required - As noted above, it is highly likely that an XML document will be transformed from one format to another at one or both ends of an exchange. The EDI translator will not completely go away but will probably be replaced by a general purpose XSLT transformer.
- XML will be appropriate for all electronic transactions and will replace traditional EDI - Traditional EDI formats offer a much more compact syntax for transmitting data than XML. They may continue to be more appropriate for large, bulk data transfers. In addition, XML is not likely to quickly replace traditional EDI where there are large infrastructure investments in traditional EDI.
- XML will make everything easier - If XML achieves its potential it will make several aspects of exchanging transactions electronically significantly easier and cheaper. However, XML deals only with the technology part of the problem. The business aspects, such as negotiating the business process, the data that needs to be exchanged and the source or destination of the data in the business application are often the most difficult parts of doing business electronically. These aspects remain the same--though there are some initiatives, notably ebXML--which may provide some help in this area.

Using XML for electronic transactions may offer significant cost advantages over traditional EDI due to:

- Off-the-shelf integration with business applications, rather than custom interfaces
- The potential to use general purpose, and hence lower cost, XML tools rather than specialized EDI software

For XML to reach this full potential, business standards must be developed for using it. Several such standards initiatives are discussed in Section 4. In addition, some remaining basic XML specifications, such as XSL and XML schemas, must be completed and approved. Finally software which implements these specifications must become widely available, and application vendors must build XML capabilities into their offerings. Many vendors are already offering significant XML tools, and this is discussed in Section 5.

4 XML E-Transaction Standards Initiatives



Time Line Progress: With the exception of completed standards in a handful of specific industries, development of XML business standards, i.e., DTDs and schemas for common business documents, is in its infancy.

There is a great deal of activity to develop business standards for using XML for electronic transactions. Several standards are being developed for use in particular industries. There is almost so much activity that it might be easier to list the industries that are *not* investigating using XML than to list those that are. While this type of activity may have been going on prior to XML coming on the scene, the advent of XML certainly seems to have accelerated it. There are also several cross-industry initiatives that seek to be universally applicable. However, despite the activity, there are only a few efforts that as yet have completed and published specifications. This section profiles a sampling of some of the major cross-industry initiatives and an overview of some of the major initiatives in education. To give an idea of the breadth of interest in XML, a small sampling of efforts in other industries is also profiled.

For a comprehensive listing of several initiatives, please refer to the article, "Extensible and More" at the XML.ORG web site. This is a source of XML information sponsored by the Organization for the Advancement of Structured Information Systems (OASIS), an industry consortium focused on SGML, HTML, and XML. The URL for the article is:
www.xml.com/pub/2000/02/23/ebiz/index.html

NOTE: The content of this section is for informational purposes only. Presence or absence of an initiative has no relationship with any PESC position on the initiative.

4.1 ANSI ASC X12

- Who: Accredited Standards Committee X12 of the American National Standards Institute
- Sponsors: American National Standards Institute
- Main Players: Major users of the X12 EDI standards, EDI translation software vendors, EDI Value Added Networks
- URL: www.x12.org
- What:
  1) Joint X12, CommerceNet, and XML/EDI Group Preliminary Report on the Representation of X12 Data Elements and Structures in XML, published August 1998. Commentary by Mike Rawlins posted on the same page. Accessible from www.disa.org/x12/x12c/x12cdocs.htm
  2) X12 Technical Report on X12-XML - An Experimental Methodology for the Representation of X12 Semantics in X12 Syntax - Published December 1999. Final published report available from X12's secretariat, the Data Interchange Standards Association. Working copies and drafts available on X12C/TG3 web page
  3) Ongoing development work aligned with the ebXML initiative

X12 began working on XML in February 1998, with the formation of an ad hoc task group with members of X12, primarily from the Communications and Controls Subcommittee (X12C), and the CommerceNet Consortium. CommerceNet is a global non-profit membership organization, founded in 1994, whose mission has been to promote and advance interoperable electronic commerce to support emerging communities of commerce. This effort was later joined by members of the XML/EDI Group, an affiliate of the Graphics Communication Association, whose mission is to promote the use of XML for electronic transactions. The task group was chartered to investigate the viability of using X12 data elements, segments, and transaction sets as a basis for using XML for electronic transactions. It was mainly concerned with determining the best way to represent X12 transaction sets in XML syntax. The final report was released in August 1998. This report proved that there was interest in using XML for electronic transactions, based on current X12 EDI standards, and provided a preliminary methodology for doing so. However, there were many deficiencies in that methodology, as noted in Rawlins' commentary.

In October of 1998, following completion of the ad hoc report, X12C took the X12-based XML work in house to develop an X12 technical report. The work was performed in the EDI Architecture task group of X12C (X12C/TG3). The goal of this report was to define a

methodology for using X12 semantics in XML syntax. The effort recognized that there is a wealth of business information contained in the X12 standards but that it was inextricably bound to the X12 syntax. The goal of the report was to devise a way to identify the business meanings, or semantics, in the X12 standards, and prescribe the best way to represent them as XML documents. The report was completed and approved in October 1999. It was published as an X12 Technical Report Type 1, which represents the work efforts of a single X12 subcommittee and does not have the consensus of a full X12 standard. It dealt only with representing an individual transaction set, and did not deal with full X12 interchanges, functional groups, and other related aspects of the X12 standards necessary for a complete XML-based solution. The work effort also developed a preliminary list of XML element names for X12 components, but at present this list has not been posted on the X12C/TG3 web site. There were plans to continue development of the methodology and involve other subcommittees of X12, but the work has been suspended with the development of the ebXML initiative. It may or may not be resumed, depending on whether or not the approach is consistent with the ebXML specifications.

X12 was invited to participate in the ebXML Work Group (see below) and a large turnout of X12 members attended the organizing meeting in November 1999. Several X12 members continue to participate, with much of the ebXML leadership coming from persons involved in X12. In light of the developments in ebXML and X12C, the Steering Committee of X12 at the October 1999 meeting formed an XML task group under its auspices to coordinate and recommend an overall strategy for XML development within X12. Based on the recommendations of that task group's first meetings at the February 2000, X12 meeting, the Steering Committee resolved that X12's XML development should continue within the framework being defined by the ebXML Work Group. There continues to be interest in most X12 subcommittees in pursuing XML development.

4.2 ebXML

- Who: The Electronic Business XML (ebXML) Work Group
- Sponsors: United Nations European Center for the Facilitation of Administration, Trade, and Commerce (UN/CEFACT, the parent body of the EDIFACT Working Group), and OASIS (the Organization for the Advancement of Structured Information Standards)
- Main Players: Wide and diverse participation, including members of CEFACT groups such as the EDIFACT Working Group, ANSI X12, major software vendors such as IBM and Sun, vendors working in XML such as CommerceOne, and trade groups from several industries.
- URL: www.ebxml.org
- What: A set of specifications for a framework to enable interoperability of XML-based business applications.

UN/CEFACT and OASIS announced formation of the ebXML Work Group in September 1999. All of the major players involved in developing XML for electronic business were invited to participate. The goal of the work group is to define an infrastructure, or framework, which will enable interoperability of XML-based e-business applications. Being a framework, it is not concerned (at least in this phase) with the development of DTDs or schemas for individual business documents. Rather, it seeks to specify consistent ways to develop, define, and transport such documents, along with specifications for supporting facilities and activities such as registries and repositories and business process modeling. The vision is that cross-industry groups such as the EDIFACT Working Group or ANSI X12, and vertical industry groups such as HL7 in Health Care, ACORD in Insurance, or the Open Applications Group in manufacturing industries, would use the specifications developed by ebXML to develop their own standards. Such standards may be developed "top down" with a consistent way to do business process modeling, with schemas or DTDs being derived from the models, or may proceed "bottom up" by directly creating schemas or DTDs. The goal is that if such standards are developed in compliance with the ebXML specifications, then systems that use them should be interoperable.

The work within ebXML falls into the following major areas, each assigned to a specific project team:

- Requirements - Determine the overall business requirements and specific requirements for the ebXML technical framework. An initial Requirements Specification has been submitted for ebXML comment, with approval expected at the May 2000 meeting.
- Architecture - Define an overall architecture for the ebXML framework. Also, deal with such related issues as XML message design.
- Transport, Routing, and Packaging - Specify the recommended means for transporting XML documents between enterprises and route them within enterprises. Also, handle security and related issues.
- Business Process - Define or adopt a business process "meta-model", i.e., a set of components or building blocks that can be used to define business processes using the Unified Modeling Language (UML), an object-oriented analysis approach.
- Core Components - Define or adopt a set of lower level components, such as name and address blocks, that can be used to define business processes and XML DTDs and schemas. The Business Process work defines components oriented towards processes, while the core components work deals more at the level of data items.
- Registry and Repository - Specify a registry for organizing and indexing XML (and other) related components, and a repository for storing them. It is envisioned that all sorts of ebXML components would be stored in these registries and repositories, along with business process models, DTDs, and schemas developed by other groups using the ebXML specifications. Trading partner profiles might also be stored here, enabling companies to consult the repository to determine the business processes and schemas supported by partner companies.

This is an ambitious work effort, with completion expected by summer of 2001. The organizing meeting was held in November 1999, and meetings follow every quarter. The project teams continue work in between meetings with weekly conference calls, work by e-mail listserv, and interim meetings. The first two meetings averaged around 150 participants, and most of these continue to be active on the project teams in between meetings.

OASIS and UN/CEFACT have plans to take the specifications produced by the work group and submit them to the appropriate bodies, such as the International Standards Organization (ISO), for adoption as full standards.

4.3 IFX

- Who: The Interactive Financial Exchange (IFX) Forum
- Sponsors: Initially organized as the "InteroperaBILL" initiative under the auspices of the National Automated Clearinghouse Association's Council on Electronic Billing and Payment
- Main Players: banks, brokerage companies, billers and technology providers, such as Checkfree, Intuit, Microsoft, AT&T, Wells Fargo, FleetBoston Financial, Security First Technologies, Just in Time Solutions, Integrion Financial Network, EDS, IBM, PaineWebber, Citibank, BankAmerica, BITS
- URL: www.ifxforum.org
- What: IFX Business Message Specification - A specification for an open and interoperable online financial services marketplace.

IFX is designed to provide a robust and scalable framework for the exchange of financial data and instructions independent of a particular network technology or computing platform. It was developed as a cooperative industry effort among major financial institutions, service providers and information technology vendors serving these institutions and customers in the small business and consumer markets. It is the result of combining the Open Financial Exchange (OFX) specification (used by Intuit in Quicken, among others), and Integrion Gold.

These are some of the main features of the IFX Specification:

- Services Provided:
  - Bank statement download
  - Credit card statement download
  - Funds transfers including recurring transfers
  - Consumer payments, including recurring payments
  - Business payments, including recurring payments
  - Brokerage and mutual fund statement download, including transaction history, current holdings, and balances.
  - Bill presentment and payment
- Provision of a Broad Range of Customer Service Providers—IFX supports communication with a broad range of Customer Service Providers (FIs), including: Banks, Brokerage houses, Insurance Companies, Merchants, Payment and Bill Processors, Financial advisors, and Government agencies
- Supports a wide variety of front end applications and clients
- Supports extensibility, reliability, security and international support
- Transport independent
- Supports batch and interactive exchanges

4.4 BizTalk

- Who: BizTalk
- Sponsor: Microsoft
- Main Players: Microsoft; wide range of participants including major software vendors such as SAP and CommerceOne, and major users such as Boeing and BP/Amoco
- URL: www.biztalk.org
- What:
  - BizTalk Framework - A set of guidelines for how to publish schemas in XML and how to use XML messages to integrate software programs
  - BizTalk Repository - A library of XML schemas conforming to the framework

Microsoft announced the formation of BizTalk in the summer of 1999. BizTalk describes itself not as a standards body, but "a community of standards users, with the goal of driving the rapid, consistent adoption of XML to enable electronic commerce and application integration." Microsoft has largely developed the BizTalk framework, with review and input by the BizTalk steering committee. Membership in the steering committee is by Microsoft invitation. The BizTalk framework is similar in intention to the ebXML framework. It covers similar things such as message transport, routing, and packaging, security, and message design rules. However, consistent with Microsoft's position that BizTalk is "not concerned with the content of documents," it does not address a consistent vocabulary for XML messages, a set of core components, or business process modeling.

The BizTalk repository is similar in function to the registry and repository being developed by OASIS at XML.ORG, which may become one of the repositories for ebXML components. Any company or organization may post schemas in the BizTalk repository so long as they follow the basic design rules defined in the BizTalk Framework. There is no attempt to order or coordinate the schemas. For example, there are over fifty purchase order related schemas, all different, with most submitted by different organizations.

Being sponsored and guided by a single dominant software vendor gives the BizTalk initiative an advantage in technical consistency and rapid development. However, the close association with Microsoft is also a liability with some potential implementers. Microsoft has recently increased

participation in the ebXML initiative, and there may yet be a chance for the ebXML and BizTalk frameworks to unify, or at least not conflict.

Microsoft is developing products to implement the BizTalk framework.

4.5 Education

This section provides an overview of XML initiatives in education known to the PESC XML Work Group at the time that this report was prepared.

4.5.1 Department of Education - Office of Student Financial Aid

- Who/What: The Office of Student Financial Assistance, a performance based organization of the Department of Education
- Sponsor: U.S. Department of Education
- URL: www.ed.gov/offices/OSFAP/ - Follow links to the Modernization Blueprint
- What: The SFA sponsored the Highway One pilot in 1999. It was a proof of concept to test various ways to collect and report data about students. FFEL loan, FDSL loan, and Pell grant data for a student were collected from multiple sources and displayed in a single, consistent view through a web browser. XML was used to format data extracted from ELMnet. As discussed in the Modernization Blueprint, XML is a key technology in SFA's future architecture for both internal application integration and all data exchanges with external entities.

4.5.2 Schools Interoperability Framework

- Sponsor: Microsoft. NOTE: The Software Information Industry Association (SIIA) has recently voted to take over management of the SIF project.
- Main Players: About 70 members as of February 2000. Major systems and software vendors such as Microsoft, Sun, IBM, Oracle, PeopleSoft, and SAP; vendors of school information systems; representatives of school districts.
- URL: www.schoolsinterop.org
- What: A specification to ensure that K-12 instructional and administrative software applications work together effectively. The specification defines standard formats for shared data (e.g., student demographics information), standard naming conventions for this shared data, and the rules of interaction among software applications.

4.5.3 IMS Global Learning Consortium

- Who: IMS Global Learning Consortium, A global consortium with members from educational, commercial, and government organizations.
- Main Players: Major systems and software vendors such as Apple, Microsoft, Sun, IBM, Oracle, PeopleSoft; vendors of school information systems such as SCT; several universities and school districts such as the University of California and Miami-Dade Community College.
- URL: www.imsproject.org
- What: Specifications for facilitating online distributed learning activities such as locating and using educational content, tracking learner progress, reporting learner performance, and exchanging student records between administrative systems.

4.5.4 ELM Resources

- Who: ELM (Education Loan Management) Resources, an alliance of student loan lenders.
- Main Players: Approximately 150 lenders, federal and state government agencies, and guarantors of student loans.
- URL: www.elmresources.com
- What: ELM Resources has created ELMnet, an ATM-scale data switch and interactive Internet client software to offer colleges uniform delivery of student loans. Currently this

service is providing remote access for schools and students, in a secure environment, to a loan service providers system in a "read only" mode. XML is used to format the data for both queries and responses. ELM Resources participated in the Department of Education's Highway One prototype.

4.6 Efforts in Other Industries

This section offers a small sampling of efforts in other industries, indicating the wide interest in using XML for electronic transactions

4.6.1 ACORD

- Who: ACORD (Agency Company Organization for Research and Development) - The insurance industry's non-profit standards developer. Initially began in 1970 serving the industry with standard printed forms, has expanded role into data integration. Other non-XML standards maintained by ACORD include AL3, a traditional EDI like standard, and ObjX and OLife object oriented standards.
- Main Players/Participants: Over 1,000 major insurance carriers and 25,000 agencies, vendors
- URL: www.acord.org
- What: ACORD XML - Designed mainly for use between insurance agencies and carriers. ACORD is developing XML-based versions of current property and casualty and life insurance standards. These are based on the IFX specification.

4.6.2 DSML

- Who: Directory Services Markup Language
- Main Players: Cisco Systems, IBM, Microsoft, Netscape, Novell, Oracle
- URL: www.dsml.org
- What: An XML-based language for network directory access, intended for network administration.

4.6.3 HL7

- Who: HL7 - Health Level Seven, an ANSI-accredited standards development organization responsible for data interchange standards for clinical and administrative data.
- Main Players: Hospital systems vendors, major systems vendors. HL7 "Benefactors" include Ernst & Young, Johnson & Johnson, MckessonHBOC, Shared Medical Systems (SMS), and the U.S. Department of Veterans Affairs
- URL: www.hl7.org
- What: The current versions of the HL7 standards are based on a traditional EDI syntax that closely resembles X12. The new version 3.0 in development will also support XML and object-oriented approaches such as CORBA

4.6.4 OAG

- Who: The Open Applications Group
- Main Players: Major ERP (Enterprise Requirements Planning) systems vendors such as SAP, PeopleSoft, Oracle, and Baan
- URL: www.openapplications.org
- What: The Open Applications Group Interface Specification (OAGIS) - A set of XML-based specifications designed to enable in-house integration of ERP and supply chain functions.

4.6.5 RosettaNet

- Who: RosettaNet, a non-profit consortium focusing on the information technology supply chain.

- Main Players: Major computer and electronics manufacturers such as COMPAQ, HP, Intel, Cisco Systems, IBM; retailers such as CompUSA, Ingram Micro, MicroAge
- URL: www.rosettanet.org
- What: XML-based standards designed primarily for use in the IT supply chain. Comprised of a data dictionary, an implementation framework that defines the infrastructure, and Partner Interface Processes (PIP) which define exchanges. The PIPs are developed using object-oriented modeling techniques.

5 Software Vendor Support for XML

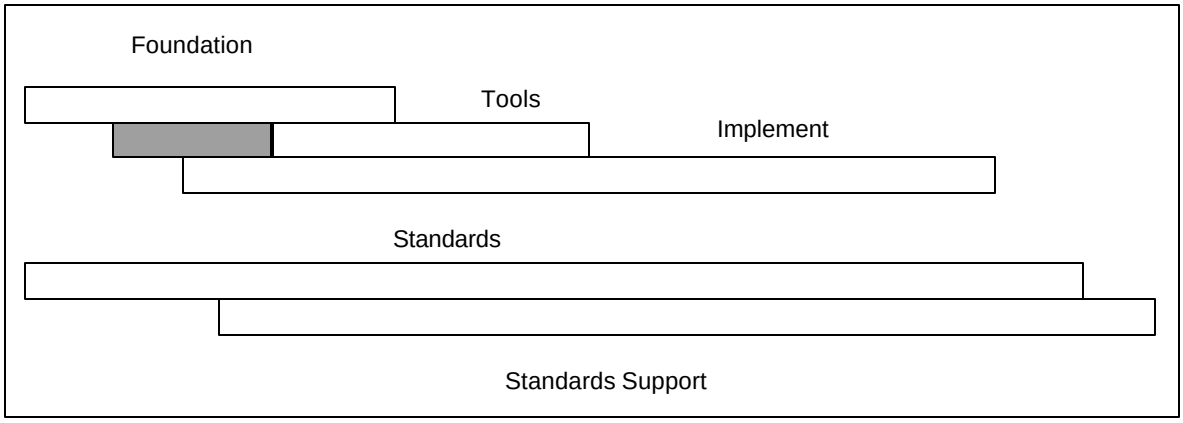
This section is intended as a representative sample of vendor support for XML. Interest in XML is so widespread that it is hard to find a vendor that does not have an XML strategy, if not product, in development. So, this survey is only a sample.

Vendors are divided into three major categories:

- Tools - Those offering tools to assist developers of applications software in building XML capabilities into their applications
- E-Commerce Software Vendors - Firms that provide e-commerce systems with XML capabilities
- Applications Vendors - Vendors of a variety of business applications

NOTE: The content of this section is for informational purposes only. Mention of or failure to mention a particular vendor should not be construed as a PESC recommendation concerning that vendor's products.

5.1 Tools Vendors



Time Line Progress: Roughly one-third complete. Support is provided for the basic approved XML specifications, but little yet for more advanced specifications and those still in development.

The vendors listed here have all made considerable investments in XML technology. They are listed here because they are major vendors, and their support and investment indicates that XML is a significant technology. A comprehensive list of tools, vendors and free XML-related software can be found on the XML Cover Pages, edited by Robin Cover and hosted by OASIS.

A common offering of many vendors is an XML "parser," which is a utility that reads XML documents and provides a means for developers to easily access and process the contents of the document.

5.1.1 IBM

URL: IBM developerWorks XML site: www.ibm.com/developer/xml

IBM dedicates a section of their web site to XML tools, articles, and news, and offers a monthly on-line XML newsletter. Some of the significant tools offered for download (currently for free) from their alphaWorks site include:

- XML parser written in Java
- XML libraries for C++
- LotusXSL, an XSL processor written in Java
- XML Generator test tool
- Visual XML Toolset

5.1.2 Microsoft

URLs:

- Microsoft's XML Developer Center - http://msdn.microsoft.com/xml/default.asp
- XML section of the Web Workshop - http://msdn.microsoft.com/workshop

Internet Explorer Version 5, as noted earlier, currently offers partial support for XML and XSL. Like IBM, Microsoft also dedicates a section of their web site to XML tools, articles, and news. Some of the significant tools offered for download (currently for free) include:

- MSXML XML parser. This is shipped standard with Internet Explorer but is frequently updated.
- XML Software Development Kit
- XML and XSL Developer's Guides
- Samples of XML processing using Visual C++
- XML Notepad
- Various XML and XSL demos

Microsoft has also recently completed initial development of its BizTalk server software, and a trial version is available free for download to developers.

5.1.3 Sun Microsystems

URL: "Java Technology and XML" section of the Sun's Java site - java.sun.com/xml

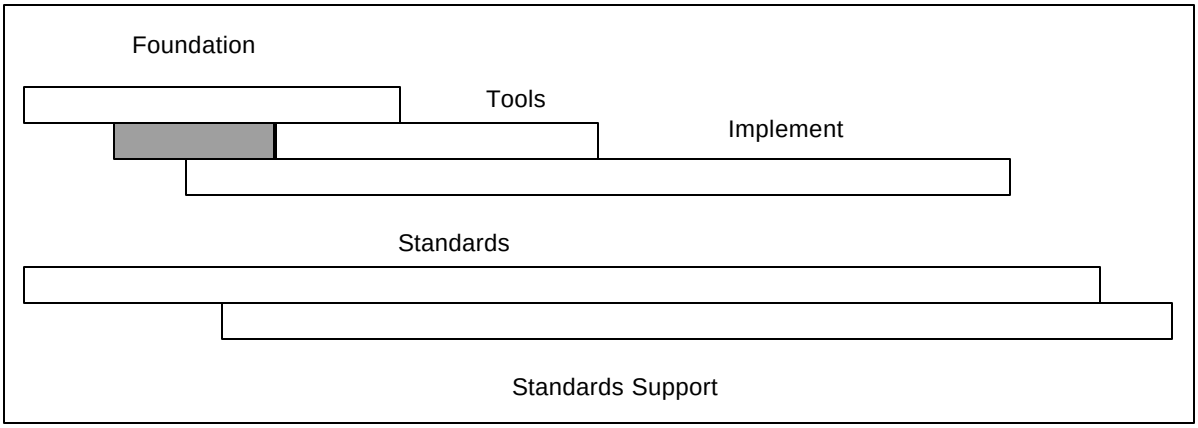
Most of Sun's XML work is related to the Java programming language, which they developed and continue to sponsor. Their XML section, like the others, has tools, white papers, and other downloads. Sun is currently developing extensions to Java that allow it to process XML. Sun states that "XML is fundamental to our plans for the next generation enterprise-computing platform, Java 2 Platform Enterprise Edition. We are using it to make Enterprise JavaBeansTM components even more portable. We also intend to make it a standard for the transmission of mission-critical enterprise data."

5.1.4 Oracle

URL: www.oracle.com/xml

Oracle, like the other vendors, has demonstrated strong support for XML for both application integration and business-to-business exchanges. Oracle has XML-enabled their entire Oracle Internet Platform, including the Oracle 8i database. They also provide the Oracle XML Developer's Kit (Oracle XDK), a set of components that facilitate the delivery and implementation of XML-based data exchanges. It is fully supported by the Oracle Worldwide Support team. The XDK, several other utilities, articles, and other items are available for download from their web site.

5.2 E-Commerce Vendors



Time Line Progress: Similar to that of tools vendors, with roughly one-third complete. Many vendors provide support for the basic approved XML specifications, although in many cases the support is fairly rudimentary. Most still do not support XML schemas, since the relevant specification has not yet been approved.

5.2.1 Sterling Commerce

URL: http://www.sterlingcommerce.com/

Sterling Commerce is a vendor of traditional EDI management and translation software, and operates a Value Added Network. Their main product families are GENTRAN, COMMERCE, CONNECTION, and VECTOR. Like most such vendors, they are building XML capabilities into existing products and seeking to position their offerings as "any-to-any" transformation utilities rather than just EDI translators. Currently XML support is offered as an option in GENTRAN:Server for Windows NT, and in the Web Suite add-on to that product.

5.2.2 Harbinger

URL: http://www.harbinger.com/

Harbinger, like Sterling Commerce, is a vendor of traditional EDI management and translation software, and operates a Value Added Network. Last year it also opened a portal site, harbinger.net, which has XML support. Harbinger has announced that it intends to build XML capabilities into its existing TrustedLink family of EDI software. It currently supports XML in the Windows desktop version. On March 13, 2000, XML support was announced in the latest version of their AS/400 offering.

5.2.3 CommerceOne

URL: http://www.commerceone.com/

CommerceOne was founded in 1994 as DistriVision, and reborn in 1997 as Commerce One. Commerce One has 600 employees, and is one of the current "high flyers" among Internet e-commerce companies. It acquired Veo Systems in 1999, which developed CBL (Common Business Language), an XML vocabulary for business documents. xCBL Version 2.0 is largely based on the UN/EDIFACT EDI standard, but deals primarily only with procurement activities and documents. XML is a key part of Commerce One's e-commerce strategy. Some of Commerce One's main products are their Buysite, MarketSite, Global Trading Web families of products and services for e-procurement and on-line auctions.

5.2.4 webMethods

URL: http://www.webmethods.com/

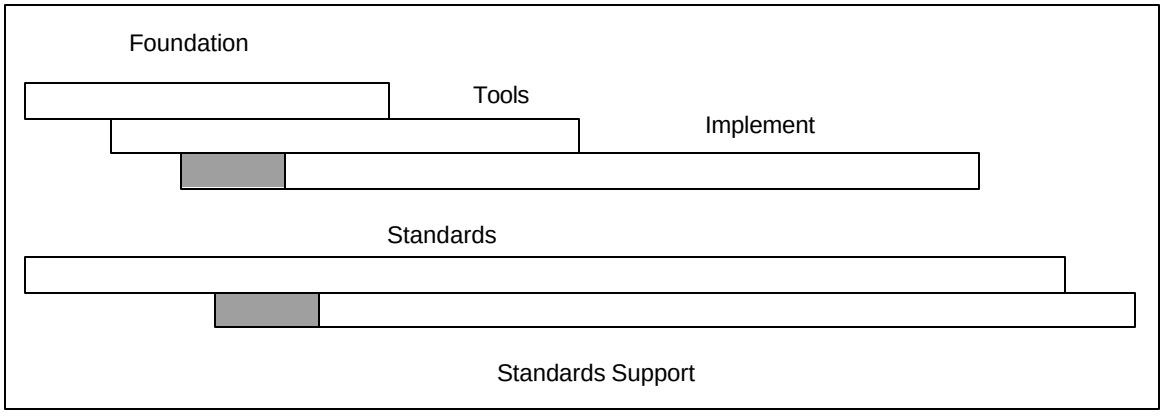
webMethods is one of a new group of firms that are offering XML-based e-commerce servers that in many ways resemble traditional EDI management and translation systems. So far, most of these are high-end solutions that only support XML, and not other formats such as X12 or EDIFACT. Their primary offering is webMethods B2B, with versions for both web portals and individual enterprises.

5.2.5 XML Solutions

URL: www.xmlsolutions.com

XML Solutions was established in 1998, and has specialized in XML-based electronic commerce. One of their primary work efforts has been to develop a methodology (dubbed XEDI) and supporting products for converting existing EDI Messages (such as X12 or EDIFACT) into XML format using information from existing EDI standards dictionaries. The approach is documented in a white paper at www.xmls.com/resources/whitepapers/X12.pdf. The approach has been presented to a few standards bodies such as ANSI ASC X12 and the ebXML Work Group but has as yet has not gained any official support from those bodies. XML Solutions' XEDI Translator is a server-based translator, implemented in Java, that takes EDI interchanges produced by an existing EDI system and re-formats them into XML for use by small to medium enterprises that do not have EDI systems.

5.3 Applications Vendors



Time Line Progress: In the initial stages. Many vendors are planning to support generic XML import and export functions but few have actually provided it as yet. Even fewer support industry XML business standards.

This section offers a high level overview of XML support from a very small sampling of application vendors who have been active with XML. In most cases and unless noted otherwise, XML is supported primarily for in-house application integration. Most vendors are working with one or more standards organizations, and have demonstrated commitment to business standards for XML.

- SAP - XML support is included in SAP's flagship R/3 product through Business Application Programming Interfaces (BAPIs). SAP is active in OAG and RosettaNet, and intends to support the specifications of those groups.
- PeopleSoft - PeopleSoft, a PESC member, in PeopleSoft 8 offers open integration with enhanced EDI/XML support. PeopleSoft also is active in OAG. It has recently joined RosettaNet and intends to support its specifications as well. XML support has been included in Student Administration 7.6 to facilitate the passing of person, class, and enrollment data to external Internet learning management systems using the IMS standards. The PeopleSoft Portal for Higher Education, in development, will also use XML.
- Intuit - Quicken currently supports importing of financial information, such as checkbook registers and credit card statements, in the OFX format (now being superseded by IFX which was discussed earlier). OFX was developed in SGML, but is compatible with XML.
- SCT Corporation - A PESC member and vendor of student information systems. SCT uses XML formatted messages to integrate student information in their back office student administrative system with the Campus Pipeline web portal.

6 Trends and Impacts

The preceding sections should provide an overview of the XML landscape. XML development in both software and business standards is rapidly moving forward. However, as has been noted, there is still quite a bit of work to be done in these areas before XML can be widely adopted for electronic transactions. Several clear trends are emerging for software vendor support. However, the area of business standards for XML is much more volatile and unclear. Several competing trends are evident, and there is not yet enough information to reliably predict how they will play out. This section addresses these major trends and their possible impacts.

6.1 Applications Support for XML

As has been noted, several major vendors of tools for application software developers are making large investments in XML and are making XML tools available. Tools from vendors such as Microsoft, IBM, and Sun are becoming increasingly sophisticated and powerful as the XML standards are approved. This trend will result in it becoming increasingly easier for application software developers to build XML import and export capabilities into their products. With a high degree of interest about XML in the trade press and increasing interest in the customer community, it will likely be common for applications vendors to provide XML support in their business applications. This is already evident with application vendors working in the postsecondary education space such as SCT and PeopleSoft, and with other major vendors such as SAP and Intuit. Other vendors of student information systems and related systems are likely to follow. With XML support built into applications, it is likely that XML will become one of the preferred means to integrate and exchange data between internal applications. For example, SCT is using XML to integrate with Campus Pipeline, and the SFA used XML in the Highway One pilot. XML will probably become much more commonplace than either the comma or tab delimited files familiar to spreadsheet and database users, or traditional "flat" files. As XML

support is built into future versions of common desktop applications such as Microsoft Office, it will likely become ubiquitous.

6.2 XML on the Web

For use on the World Wide Web, XML will probably gradually replace HTML as necessary standards are approved and as browsers and web publishing tools (or XSL authoring tools), such as Microsoft FrontPage, support it. As this happens, new applications will emerge that are either only speculated about now, or not even imagined. For example, new types of search engines will emerge that can scan XML documents with keywords contained in XML tags. Web-based course catalogs could have keywords embedded in XML tags to aid searches for specific course content. If the World Wide Web Consortium's vision for XML is realized, the web will become a much different place.

6.3 XML Complexity

There is one contrary trend, however, which may have a quite different impact. A history of computing (and other areas, too, for that matter) shows us that simple, easy solutions generally gain wider acceptance than complex and difficult ones. For example, the relatively simple TCP/IP protocol of the Internet has become the basis of the world wide computer network rather the complicated OSI protocols (Open Systems Interconnect) developed in the 1980s to be the basis for the world wide computer network. Similarly, the Internet's SMTP mail protocol has made OSI's X.400 almost vanish. One of the great strengths of HTML has been its relative simplicity. Almost anyone can use it, even grade school students. Basic XML is fairly simple. However, as more complicated features such as XSL are built onto it in order to provide a complete solution, it is becoming much more complex. If the W3C designers do not attempt to hold this complexity somewhat in check, it may become increasingly difficult for software vendors to provide low cost, easy to use support for XML. This is a trend that bears watching. It may be measured by such things as the rate of introduction of and number of XSL authoring tools, and the rate at which application vendors support XML schemas. If vendor support lags approval of the XML specifications by months or even a couple of years, this trend will be confirmed. If confirmed, the trend toward XML ubiquity will be severely curtailed.

6.4 XML Business Standards

As noted, the picture for XML business standards is not quite so clear. After XML was first approved in 1998, there was a trend to promote individual, proprietary implementations on the basis that XML's flexibility made standards unnecessary. However, this trend quickly died out for the most part. Perhaps learning from past experience with traditional EDI, most companies have been waiting for business standards to emerge before embracing XML for electronic transactions. 1999 saw trends toward vendor- promoted standards (such as CommerceOne's CBL and Ariba's cXML), and standards in specific industries. The vendor-promoted standards have largely focused on procurement and garnered wide interest when announced. However, after the initial interest died down and most of the early adopters had made their commitments, acceptance of these standards seems to have stalled. Users seem to want a stamp of legitimacy and support beyond what can be offered by a single vendor. Even Commerce One, whose CBL is perhaps the most prominent vendor-promoted standard, seems to recognize this and has committed considerable resources to participating in the ebXML initiative.

The trend in vertical industries continues unabated, however. As noted earlier, there have always been initiatives to develop electronic transaction standards in specific industries and for specific niches. However, most of them now seem to be using XML, and there seem to be more such initiatives than there were prior to XML coming on the scene. These initiatives are using XML for

both external transactions and internal application integration. The Schools Interoperability Framework in K-12 is a good example of the latter type of initiative. If the trend continues, we may expect to see a similar initiative develop for integrating applications in the college and university environment. These initiatives will continue to have an important role. Framework initiatives such as BizTalk and ebXML expect that most of the actual document DTD or schema development will be performed in vertical industry groups, only consistent with the guidelines defined by the frameworks.

However, for many essential business functions, such as purchasing, invoicing, and shipping, there are as yet no prominent cross-industry standards efforts. ANSI X12 and the EDIFACT Working Group certainly have a library of documents that could be easily converted to XML, but that work is suspended pending development of the ebXML framework. In addition to this lack of cross-industry documents, ebXML currently is the only initiative developing a standards-based cross-industry framework, and it is not expected to begin delivering any specifications until May 2000. Many ebXML participants believe that it is unlikely for another such cross-industry initiative to emerge. Microsoft's BizTalk framework is not as comprehensive and as yet has only tentatively been embraced. In this vacuum, many users are reluctant to implement XML. So, success of the ebXML initiative is important to developing cross-industry XML business standards and a cross-industry, non-proprietary framework. Without a succesful cross-industry initiative, there will be competing XML standards for common cross-industry documents such as purchase orders and invoices, and competing standards for transporting and processing them.

If ebXML succeeds, it should enable widespread implementation of XML for electronic transactions by providing a single, consistent framework. However, if it fails, the trends toward development in vertical industries may fill the vacuum, but also lead to further problems. For example, the Schools Interoperability Framework could extend its scope beyond K-12 into postsecondary education. In other industries, the OAG initiative, which primarily focuses on application integration within an enterprise, is already extending its focus to exchanges between enterprises. In addition, RosettaNet is reported to have plans to extend its focus into other industries beyond just the IT supply chain. If this trend continues, there will be competing standards and users will be left with a different kind of problem, i.e., which standard to implement rather than having no standard to implement. The impact on the PESC community may not be as severe as in other industries, since at present there are few competing XML business initiatives in this area. However, those involved in procurement, and to a lesser extent, financial transactions, may be faced with having to support multiple standards.

This scenario could still come to pass in the event that ebXML is either behind schedule or software vendors do not adopt its specifications. Since X12 and EDIFACT have committed to ebXML, their success in providing cross-industry DTDs and schemas is dependent on ebXML's success. If ebXML fails and X12 and EDIFACT must proceed on their own, the time lost in waiting for ebXML may further aggravate the proliferation of competing standards. It may then take several more years for the XML business community to feel enough pain from implementing competing standards that it again undertakes another attempt at a single, all encompassing standard.

In the broad picture, it will probably take two or three years to determine which of these XML business standards trends will dominate. The progress of the ebXML initiative bears close monitoring as well as vendor support of its specifications as they are released. Progress of user communities in embracing vertical industry standards such as RosettaNet in industries outside of their origins should also be monitored as well as adoption of internal integration standards such as OAG for inter-enterprise exchanges.

6.5 Registries and Repositories

Another trend, related to XML business standards, has been the development of registries and repositories for XML DTDs and schemas. Repositories can be thought of as libraries of XML components, with registries serving the function of catalogs, allowing the DTDs and schemas to be publicized and shared among users. BizTalk already has a repository on-line, and OASIS is developing XML.ORG repository. The ebXML Work Group is developing a specification for a network of linked registries and repositories. At present, there are only these two major competing repositories, and if no others surface, then it may be fairly easy to locate desired DTDs or schemas. However, if several others develop and they are not linked with each other, finding and sharing DTDs and schemas could become more difficult than at present.

6.6 Summary and Impacts

For the PESC community, perhaps the most notable aspect of these trends at present is that there are relatively few XML business standards either in place or in development. However, nature (and the marketplace) abhors a vacuum. It would be unusual if one of these trends did not manifest itself with the development of XML standards in the postsecondary education space. Although the development of vendor-sponsored XML business standards seems to have slowed generally, there is certainly a chance that, lacking any available alternative, a vendor may propose a proprietary standard either for any of the common electronic transactions currently in use or for on-campus integration. There are also a few approaches such as XML Solutions, in addition to X12C's preliminary work, for converting X12 transaction sets to XML messages. If one of these approaches gains acceptance in the market before X12 develops a full solution, it could start to appear in the PESC space. As noted, other standard initiatives such as the Schools Interoperability Framework or RosettaNet may expand their scope into the PESC space. Clearly, coordination is needed to prevent competing standards from emerging.

Finally, even if XML becomes widely adopted for electronic transactions, it is likely that traditional EDI will persist. There is a significant installed base and heavy investment in traditional EDI technology. For many types of applications, the benefits of moving to XML (mainly easier use by small to medium enterprises), may not justify the conversion costs for some time to come. In addition, relatively small, simple, near real-time, commonly used transactions, such as those related to procurement, will probably be the first to become widely accepted. More complex, batch-oriented transactions, such as student transcripts or health care claims, may remain in X12 for years to come.

7 Recommendations

Based on the analysis presented here, the PESC XML Work Group offers recommendations for three constituencies: PESC regarding standards development activities, developers regarding XML support, and users regarding implementation strategies.

.

7.1 PESC XML Standards Development

As noted, the two greatest problems concerning business standards for XML are:

(1) A lack of standards, leading to a multitude of differing proprietary implementations
(2) Competing, overlapping standards

The XML Work Group recommends that PESC take action to alleviate these problems by promoting development of a single, unified approach to XML business standards for the PESC community. To do this, PESC must:

- Recognize and embrace the emergence of XML as an important technology for electronic exchange.
- Recognize the need to develop XML business standards, and work aggressively with the appropriate bodies to plan for development of such standards. Such standards should encompass not only transactions between higher education institutions/organizations, but transactions and integration within them. Several aspects of that work include:
  - Formulate a position on support for frameworks such as ebXML and BizTalk
  - Since business process models and information models are increasingly being used as a basis for standards development, work with appropriate bodies to insure that this work is being performed. Existing and planned PESC work on data dictionaries is an excellent basis for this work. This is consistent with the position of most frameworks that industry-specific XML element names and schemas should be constructed using industry data dictionaries. PESC may wish to offer its website as an XML repository to serve as a resource for the higher education community.
  - Identify the business processes and X12 EDI transactions that might be most appropriate for moving into XML. Those with a small installed base and relative transaction simplicity might be the most appropriate to convert first. Student transcripts might be the last for consideration due to an established user base and greater complexity.
  - Monitor and use PESC influence to coordinate work of the various mature organizations developing XML business standards for the PESC community, with the goal of preventing overlapping and competing efforts.
  - Provide guidance to emerging and immature XML initiatives.
  - Monitor and encourage XML advocates to incorporate the ANSI qualities of consensus development and public access to all standardization efforts.

In the past, PESC has served primarily to coordinate standards development and promote use of standards. However, in light of current developments in XML and the absence of any single organization responsible for developing XML business standards for the postsecondary community, the XML Work Group recommends that PESC take a more active role in and responsibility for development of such standards.

An approach for accomplishing this, the Work Group proposes, is to form a group of community stakeholders in XML development, with PESC providing any needed XML expertise. This group would focus on the bulleted items above. Coordination with the PESC data definition repository effort is recommended. With a number of XML implementations already taking place in higher education, the Work Group encourages PESC to move quickly to ensure the leadership and representation necessary to develop community standards.

7.2 Developers

By now, most applications software vendors are at least aware of XML, if not already forming strategies for supporting it. As previously noted, it is likely to become at least common import/export format, if not the basis for electronic transactions. Developers would be well served to begin researching the available development tools and make plans for providing at least basic import/export features. Basic import and export features should be built with a flexible approach so that XML element names may easily be changed to match industry standard names as these are developed. Developers are advised against creating their own XML vocabularies, DTDs, or schemas for electronic transactions. They are also advised to wait until standards for intra-campus integration are put in place before developing XML-based campus integration frameworks.

Large universities and others with internal development efforts would likewise be well served to begin researching and familiarizing technical staff with XML. Pioneers may wish to begin using XML for selected internal projects and generic import/export capabilities. However, the Work

Group recommends refraining from developing external interfaces for electronic transactions until standards are put in place.

7.3 Users

With the lack of XML business standards and the still relative immaturity of the base XML standards and technology, the Work Group recommends against general deployment of XML for electronic transactions at this time. Deployment in some niche areas, such as learning activities using the IMS standards, may be appropriate where no similar EDI standards exist. Users are advised to become familiar with XML technology and stay abreast of the offerings of software vendors. At present, a good guess is that XML may be ready for general deployment in two or three years, but this is only a guess. However, there are several clear dependencies. The time line used throughout this document can be used as guidance. Users may monitor progress on the time line to estimate when they should begin deployment.

For users considering further investments in traditional EDI technology, if the ROI can be attained within two to three years then the investment is certainly justified. Traditional EDI will likely endure and co-exist with XML electronic transactions for some time to come. If the ROI takes longer than three years the investment may still be justified. However, XML technology is rapidly evolving, and there is a slight chance that traditional EDI could become obsolete in as soon as five years. So, the risk of not achieving the desired return increases with more extended payback periods.

As noted, traditional EDI may still be more appropriate for several years for complex standard transactions such as student transcripts. The Work Group encourages institutions currently considering moving to electronic transcripts to adopt the current X12 transactions and not wait for XML messages to be developed.

8 For More Information

Tim Pavlick - PESC XML Work Group Chair - tpavlick@kpmg.com
Lysbeth Bainbridge, PESC Executive Director - Bainbridge@standardscouncil.org
Mike Rawlins, Technical Consultant to the Work Group - rawlins@metronet.com