RESPONSE TO THE DEPARTMENT OF EDUCATION ASSESSMENT TECHNOLOGY STANDARDS REQUEST FOR INFORMATION

Pacific Metrics submits this response to the Department of Education's Assessment Technology Standards Request for Information (RFI) as published in the Federal Register Vol. 75, No. 243. We have responded to selected questions in Section 3.2—Questions About Assessment Technology Standards.

Section 3.2—Questions About Assessment Technology Standards

Technological Questions Regarding Assessment Technology Standards

3.2.8 Interoperable Assessment Instruments. What techniques, such as educational markup or assessment markup languages (see also http://en.wikipedia.org/wiki/Markup_language), exist to describe, package, exchange, and deliver interoperable assessments? How do technology standards include assessments in packaged or structured formats? How can technology standards enable interoperable use with resources for learning content? How can technology standards permit assessment instruments and items to be exchanged between and used by different assessment technology systems?

Most technology standards work as it applies to assessment interoperability has revolved around the definition of content formats such as QTI, APIP, SCORM and etc. and there has not been wide spread adoption of any one format.

Electronic exchangeable item content is only part of the interoperability problem which includes but is not limited to:

- Electronic data exchange through format adoption or transformation
- Registration and Discovery of Services
- Authentication

Electronic data exchange through format adoption or transformation

There are two (2) ways solve the problem of exchanging data between disparate systems:

- Adopting a single standard as the exchange format
- Employing mapping and transformation

Adopting a single a standard as the exchange format for item content seems to be the more popular of the two (2) approaches which unfortunately does not address other data interoperability concerns such as student information. While this technique will certainly work if wide spread adoption occurs it is very narrow in scope and does not scale.

The alternative is to define a technology standard that is comprised of a master schema for all information object types including but not limited to:

- Item content
- Student content
- Education content
- Standards content

Each customer (state, district, etc.) and vendor could then decide to support the full schema or a subset of it.

Information trading partners would implement object mappings that would be used to transform data objects guaranteeing the resultant could be consumed. This approach is prevalent and has been very successful in the Electronic Data Interchange (EDI) community for many years.

The technology standards that could be employed to facilitate this model are:

- Extensible Markup Language (XML)
- XML Schema Definition (XSD)
- Extensible Stylesheet Language Transformations (XSLT)

Registration and Discovery of Services

Not all customers and vendors will necessarily support all interoperability services so there needs to be a mechanism where services can be registered and programmatically discovered. For example, an assessment platform may need item types that are distributed across multiple content systems. The assessment platform should be able to go to a universal registry and ask for services that can provide the given item types. The assessment platform would then invoke the appropriate services never really knowing who the publisher is or where they reside.

The technology standards that could be employed to facilitate this model are:

- Universal Description, Discovery and Integration (UDDI)
- Web Services
- Web Services Description Language (WSDL)
- Simple Object Access Protocol (SOAP)
- Web Services Security (WSS)

Authentication

When a user authenticates into an assessment session it should not be required that the validation occur within the assessment platform. Customers should be able to provide their users with single sign on capabilities independent of what system the user is authenticating into so they can use the same credentials and permissions do not need to be propagated.

Maintaining multiple copies of user profiles across vendors and customers is inefficient, adds complexity, and is subject to security issues.

Some technology standards that could be employed to facilitate this model are:

- Security Assertion Markup Language (SAML)
- OpenID
- WS-Federation

3.2.9 Assessment Protection. For this RFI, ''Assessment Protection'' means keeping assessment instruments and items sufficiently controlled to ensure that their application yields valid results. (See also paragraph below,' Results Validity.'') When assessment instruments or content are re-used or shared across organizations or publicly, are there capabilities or strategies in the technology standards to assist in item or instrument protection? What mechanisms or processes exist to ensure that assessment results are accurate and free from tampering? Do examples exist of public or semi-public assessment repositories that can provide valid tests or assessments while still sharing assessment items broadly?

All item content and student scores should always be maintained behind the firewall. If this information is shared it should be done using secure Web Services. WS-Security is a member of the WS-* family of web service specifications and was published by OASIS. The protocol specifies how integrity and confidentiality can be enforced on messages and allows the communication of various security token formats, such as SAML, Kerberos, and X.509. Its main focus is the use of XML Signature and XML Encryption to provide end-to-end security.

3.2.10 Security and Access. In what ways do technology standards provide for core security issues, such as access logging, encryption, access levels, and inter-system single-sign-on capabilities (i.e., one login for systems managed by different organizations)?

There are really no technology standards for access logging which is generally implemented in the service framework by the application or in the authentication component. Some vertical applications such as Content Management Systems (CMS), Document Management Systems (DMS), etc. provide access logging at a much finer granularity but the implementations are typically proprietary.

Some common encryption technology standards are:

- Secure Socket Layer (SSL)
- XML Signature
- XML Encryption
- MD5
- SHA-1

There are really no technology standards for supporting access levels which are generally implemented in the database and application.

Some common single sign on technology standards are:

- Security Assertion Markup Language (SAML)
- OpenID
- WS-Federation

3.2.11 Results Validity. For this RFI, ''Results Validity'' means protecting the statistical validity and reliability of assessment instruments and items. How can interoperable instruments be managed to ensure they are administered in a way that ensures valid results? Are solutions regarding assurance or management of validity appropriate for inclusion in technology standards, or should they be addressed by the communities that would use the technology standards to develop specific assessments?

Solutions regarding assurance or management of validity should not be included in technology standards.

3.2.12 Results Capture. How can technology standards accurately link individual learners, their assessment results, the systems where they take their assessments, and the systems where they view their results? How do technology standards accurately make these linkages when assessments, content, and other data reside across numerous, distinct learning and curriculum management systems, sometimes maintained by different organizations?

There is no technology standard that directly addresses this issue; however, there are applicable models of how the problem should be solved. ZooKeeper is a centralized service that acts as a registration proxy for distributed services. When a given application needs a particular service it requests the service from Zookeeper which returns a pointer to where the service lives. The client Zookeeper wrapper takes that location information and performs a remote invocation of the service. This model facilitates a seamless distribution of services without burdening the clients of the services.

A centralized service would register activities as reported by different applications pertaining to the student, who would be assigned a Global Unique Identifier (GUID). So if a student wanted to review assessment results the client application would first make a request to the centralized service asking for remote services that have registered as providers of assessment results. The client would then present metadata information pertaining existing assessment results to the student. The student would then select the appropriate assessment results and the client would fetch them using the remote service.

3.2.13 Results Privacy. How do technology standards enable assessment results for individual learners to be kept private, especially as assessments results are transferred across numerous, distinct learning systems? How can such results best be shared securely over distributed set of systems managed by independent organizations that reauthorized to receive the data, while still maintaining privacy from unauthorized access?

All assessment results should always be maintained behind the firewall. If this information is shared it should be done using secure Web Services. WS-Security is a member of the WS-* family of web service specifications and was published by OASIS. The protocol specifies how integrity and confidentiality can be enforced on messages and allows the communication of

various security token formats, such as SAML, Kerberos, and X.509. Its main focus is the use of XML Signature and XML Encryption to provide end-to-end security.

3.2.14 Anonymization. Do technology standards or technologies permit or enable anonymization of assessment results for research or data exchange and reporting? How do various technology standards accomplish these tasks? For example, where a number of students take a test, can their answers be anonymized (through aggregation or other techniques) and shared with researchers to examine factors related to the assessment (e.g., instructional inputs, curriculum, materials, validity of the instrument itself) without revealing the identity of the learners? Is this an area where technology standards can help?

Anonymization is accomplished by applying normalization to the data model. The assessment results should be stored separately and without any student identifying information. A linking table containing assessment results and student identifiers can be used for referential purposes.This way assessment results can be pulled from the appropriate tables without having to perform any filtering.If non-identifying student information such as grade or age is needed then the linking table will provide the joining mechanism so that the data can be efficiently queried and fetched.

This can be accomplished using a Relational Database Management System (RDBM) such Mysql, Oracle, SQLServer, and etc.

3.2.15 Scoring and Analysis of Results. How can technology standards be used for the scoring, capture, recording, analysis or evaluation of assessment results?

Scoring is typically implemented as an automated or manual process and is generally a custom solution. Like assessment results, scoring results are persisted using a Relational Database Management System (RDBM). Once in the database the scores can be efficiently queried and fetched by other applications such as an analysis engine. In many cases the final destination for the assessment results and scoring information is a Data Warehouse where longitudinal analysis and reporting can be performed.

3.2.15.1 Results Aggregation and Reporting. How can technology standards enable assessment results to be aggregated into statistical or other groupings? How can technology standards provide capabilities for results (aggregated or raw) to be reported across multiple technology systems? For example, if a learner takes an assessment in one system, but the results are to be displayed in another, how do technology standards address transferring results across those systems? How do technology standards address aggregation of results for number of learners who are assessed in one system and whose results redisplayed in yet another technology system? Can anonymization controls be included with aggregation and reporting solutions to ensure individual data privacy and protection (see also 3.2.14 above).

See 3.2.12.

3.2.16 Sequencing. How do technology standards enable assessment items stored within an assessment instrument to be sequenced for appropriate administration, when the assessment consists of more than a single linear sequence of items? For example, how do technology standards address computer-adaptive assessments? How are the logic rules that define such sequencing embedded within a technology standard?

The heuristics used to construct a test abstraction generally inject item sequencing. Most assessment administrations will consist of multiple forms per curriculum where a form is defined to be a certain permutation of items. There is no technology standard that addresses the heuristics for test construction which should be implemented as a Rules Engine for re-use purposes.

Ideally a technology standard should be created that defines test Rule Sets so that vendors can implement Rule Engines that can consume the Rule Sets.

This would solve the problem of how to aggregate items into a test abstraction once they have been fetched from a item content repository.

3.2.17 Computer-Driven scoring. How do technology standards permit, enable, or limit the ability to integrate computer-driven scoring systems, in particular those using ''artificial intelligence,'' Bayesian analysis, or other techniques beyond traditional bubble fill scoring?

Automated scoring algorithms use a myriad of computing techniques and are generally implemented as custom solutions. "Artificially intelligent" systems generally imply artificial learning technologies and are not the preferred technique or terminology. This is because these platforms are by definition unstable in the scoring environment.

The ideal situation would be to develop a technology standard that defines how the results of specific item types should be scored in the form of Rule Sets that a scoring Rules Engine could consume.

3.2.18 Formative, Interim, and Summative Assessments. What technology and technology standards exist that support formative, interim, and summative assessments? What technology standards support nontraditional assessment methods, such as evidence, competency, and observation-based models?

Some assessment vendors provide formative, interim, and summative offerings generally implemented as web-based solutions. There are no technology standards that define how the solutions should be implemented. Technology standards common to most vendors revolve around specific delivery and persistence.

Technology standards commonly used in delivery are:

- Hypertext Transfer Protocol (HTTP)

- Hypertext Markup Language (HTML)
- Secure Socket Layer (SSL)
- Web Browser
- JavaScript
- Java Platform Enterprise Edition (JavaEE)
- Hypertext Preprocessor (PHP)

Most assessment vendors use Relational Database Management System (RDBM) to persist data.

If non-traditional assessment methodologies are to be implemented as web-based solutions, then the technology standards would be similar to that of formative, interim, and summative assessments.

3.2.20 Repositories. What technology standards-based assessment instruments, questions, or item banks (or repositories and learning management systems) are used to manage and deliver assessments?

Most content development environments are custom solutions that are typically backed by a Relational Database Management System (RDBM). Items are edited and published in the content development environment and then pushed to a Relational Database Management System (RDBM) that the assessment delivery engines use.

The ideal situation would be to develop a technology standard that defines:

- What item markup (XML) and metadata schemas (XSD) must be supported
- What editing, deletion, versioning, and publishing features content development environments must support
- What data exchange services (Web Services) content development environments must support
- How external resources are linked to items

This technology standard would allow for the independent publishing and consumption of item content.

3.2.21 Content Lifecycle. How can technology standards be employed to support an assessment content lifecycle (creation, storage, edit, deletion, versioning, etc.)?

There are many commercial and open source Content Management Systems (CMS) and there are a couple of technical specifications. Unfortunately the typical functionality implemented in a CMS would not satisfy the requirements of an item bank.

A technical specification needs to be created to specifically define the functionality needed to support the full life cycle of item content.

Most item banks are custom solutions that typically implement the functionality needed to author and publish items.

See 3.2.21.

3.2.22 Interfaces and Services. What interoperability specifications for application program interfaces (APIs) or Web services interfaces to assessment management, delivery and tracking systems have been developed? How are they organized? What are the best practices related to their design and usage? How broadly have they been adopted, and what are the lessons learned from those who have designed or implemented them?

Most assessment vendors have not adopted Web Services and in most cases where they have been implemented it has been done in the context of the enterprise.

Web Services are the de facto standard for secured data exchange in most other sectors such as financial, energy, and etc.

Web Services have gained wide spread adoption and are typically the plumbing to most Service Oriented Architectures (SOA).

There are many commercial and open source frameworks such as Apache CXF that have precedence in the market and are well supported.

There are many Web Services publications and web articles that address best practices for the design and implementation of Web Services.

Web Services is one of the technology standards that the educational sector should consider as an integral interoperability tool.

3.2.23 Internal Transparency and Ease of Use. Are there technology standards and communication protocol implementations that are "human readable?" What are the benefits and risks of "human readable" technology standards? Some technology standards are not comprehensible without tools to unpack, decode, or otherwise interpret the implementation data resulting from use of the technology standard. Other technology standards, such as HTML, RTF and XML, are largely readable by a reasonably sophisticated technical user. RESTful-designed Web services are often specifically intended to be readable by, and even intuitive to, such users as well. We ask commenter's to consider the extent to which various technology standards possess native' human readability" and comprehensibility.

There are many types of markup languages and most have been designed for very specific purposes. For example, HTML is a markup language that was specifically designed for presentation purposes and to work in conjunction with HTTP. It is not a good candidate for modeling data objects such as items.

XML is the most common markup language used for modeling data objects for the following reasons:

- There are many commercial and open source XML parsers available
- XML can be validated using an XSD
- XML is the object mapping markup used in the Java Bean framework
- XML can easily transformed using XSLT
- XML supports typing using an XSD
- XML is the markup used by the SOAP protocol
- XML is the markup used by WSDLs to describe Web Services

There are other markup languages that have very vertical uses such as JSON. JSON is very light weight when compared to XML and is commonly the wire protocol for REST services used in conjunction with JavaScript and Ajax.

Most of the common markup languages are human readable and the required technical expertise of the reader depends on the complexity of markup representation.

3.2.24 Discovery and Search. How is the discovery of items or instruments (or other elements) handled within a technology standard or technology? For example, are there search APIs that are provided to permit a search? How are metadata exposed for discovery by search engines or others?

Most item content is generally maintained behind the firewall which makes it impossible for crawlers to access any information pertaining to it. Access to item content and metadata needs to be performed in a discretionary and secure manner. Rather than publishing data, it is more practical to publish secure Web Services to a Universal Description, Discovery and Integration Registry (UDDI) that can be discovered. This allows the owner of the item content to expose what information is appropriate. The UDDI allows content providers to register Web Services and content users to locate and consume them. The secure Web Services permits access to public information from behind the firewall and delivers it back in a Extensible Markup Language (XML) format. XML is a machine readable format that most Application Programming Interfaces (APIS) can consume. The Web Service would deliver back the metadata description or schema of the item which will vary among content providers. Rather than attempting to agree on one metadata structure it is more practical to provide Extensible Stylesheet Language Transformations (XSLT) that can be used to convert item content from one form to another. This approach is prevalent and has been very successful in the Electronic Data Interchange (EDI) community for many years.

So if a given entity (i.e., person, application, etc.) were searching for say K-12 Math items the steps in the aforementioned process would be:

- Locate the UDDI
- Query the UDDI for registered item search Web Services
- Invoke the Web Service with parameters as defined by the Web Service Description Language (WSDL)
- Receive the item content and translate using XSLT into a consumable import format

If a central or local search repository is needed to support advanced queries, an indexer can populate the item corpus by periodically extracting item content using Web Services, indexing it, and updating the repository.

3.2.25 Metadata. What kinds of metadata about assessments (i.e., information describing assessments) are permitted to be stored within technology standards or technologies? How do technology standards accommodate structured data (such as new State curriculum standards) that were not anticipated when the technology standard was designed? How are metadata describing unstructured (such as free-text input) and semi-structure data incorporated within assessment technology standards?

Assessment metadata should be represented using XML and should always reference an XML schema (XSD) which can be used to validate it. For a myriad of reasons assessment metadata will vary across states, districts, etc. XML is designed to handle structured data such as content areas and the standards by which they are measured. More importantly it is extensible and an evolving XSD be can versioned to reflect the existing state of the schema. XML elements can be typed through the XSD providing a level of integrity and validation most Markup Languages don't support. Unstructured and semi-structured data would be defined as text elements and could provide structured information about the content using attributes. XML is machine readable and can easily be consumed by assessment applications. XML has been widely adopted by the technology community and many different business sectors have defined standard XSDs to facilitate data interchange.

3.2.26 *Recommendation, Rating, and Review*. Do technology standards or technologies permit rating, review, or recommendations to be incorporated within an item, instrument, or other element? If so, in what ways? How are conflicting ratings handled? Do technology standards or technologies permit ''reviews of reviews'' (e.g.,' thumbs up/down'' or ''Rate this review 1–5'')? Is the rating or review system centralized, or are multiple analyses of the rating data permitted by distributed participants?

There is no technology standard for defining rating strategies, models, and reconciliation.

3.2.27 Content and Media Diversity. What types of diverse content types and forms of assessment content exist that extend beyond traditional paper-based assessments translated to an electronic delivery medium? We are interested in learning more about electronic delivery and

interaction media, such as performance-based assessments, games, virtual worlds, mobile devices, and simulations.

Online assessment data has very quickly moved from simple HyperText Markup Language (HTML) forms to a fusion of media types such as Adobe Flash, audio, video, etc. As the media types become more complicated so do the assessment delivery platforms. Technology stacks designed for common web access are no longer suited for the increasing complexity associated with new media types and the changing behaviors of assessment strategies. In addition it can no longer be assumed that the client will be a browser and the hardware is a desktop machine. Feedback generated by strategy-based assessments is much different from conventional assessments and has ranging implications regarding the structure of the information presented and collected. Presentation must be decoupled from the delivery layer and services should only be concerned with delivering data in the most efficient manner possible. Persistence models and content development systems must change to accommodate the varying requirements and structures of new media types. It is no longer sufficient to rely strictly on Relation Database Management Systems (RDBMS) as the de facto storage technology. Other persistence technologies such as NoSQL are more suited for the richer media types. The Hypertext Transfer Protocol is fine for conventional web access but due to its stateless and heavy weight nature is not suited for strategies requiring highly interactive strategies such as virtual worlds or gaming.

Questions regarding this submission should be directed to:

Stella Gibbs
Executive Vice President
Pacific Metrics Corporation
585 Cannery Row, Suite 201
Monterey CA 93940
Tel: 831-646-6402
Email: sgibbs@pacificmetrics.com
www.pacificmetrics.com