



Recommendation for the Adoption of Single File Schema for PESC Standards

Prepared by the PESC Technical Advisory Board

Introduction

PESC standards use the W3C Schema Language to specify the structure of interchange messages. These schemas allow for the precise specification of data content and element ordering. PESC has recognized two types of schema objects: schema elements and types that are specific to a particular application domain (called a sector) and those that are generally applicable to all applications (called core main). As new standards were developed, components were added to these schemas and imported into the top level schema. This approach of using an import dependent schema causes several problems.

1. **Validation Resources.** Validation of an instance document by the schema requires more memory and processing capability with each new release. While processing capacity and memory for information systems is increasing, the TAB has received reports of some validations taking significant time and requiring significant computer resources.
2. **Library Version Assembly.** Someone wishing to validate an instance document must also determine exactly which versions of each library are required and place them in the appropriate location for validation. If multiple versions of the libraries are kept, it is possible that a developer may use the wrong version of a component and thus allow defects to propagate through the processing software for creating messages.
3. **Version Management.** PESC has to manage many versions of sector and core libraries. Since a standard explicitly names its import libraries and their versions, the library schemas must be maintained until the standard is upgraded to a more recent library. This also requires that the CCB tracks the versions used by each standard to know when an older version can be deprecated.
4. **Change Management.** Multiple versions of libraries discourage using the latest version since it is difficult to know which components have changed from one version of a library to another. This occurs because changes to a library version may include changes from several standards. This makes moving to a later library risky and requires a labor intensive mapping between the old and the new libraries.
5. **Component Identification.** It is difficult to determine exactly which components are actually used by a standard. To determine the components in a standard, one needs to walk through every “type”, “ref”, and embedded type to discover all components used in a standard. While this process can be automated, this requires programming that has not been generally available to the PESC community.

Single File Schema Approach

The single file schema approach keeps all components (elements, types, and groups) for a standard in a single file. This schema can be used to validate application messages and to create messages. This approach resolves the issues listed above with multi-file schemas since the schema is completely self-contained and has no dependencies on any other schema:

1. **Validation Resources.** Since the schema is considerably smaller and is only in one document, memory and processing requirement should be significantly

reduced. The table below shows the reduction in size of some standards that have been converted:

Schema	Standalone	Combined	Savings	Percent of Combined
ETSR	5407	12101	6694	44.68%
Recruitment and Enrollment	6461	12215	5754	52.89%
Admission Application	6289	11552	5263	54.44%
IPED 4 Year graduation	198	11266	11068	1.76%
High School Transcript	4715	11980	7265	39.36%

2. **Library Version Assembly.** There is no possibility of mixing up versions of sector libraries since there are no libraries to import.
3. **Version Management.** Once the single file has been created, there is no longer any need to maintain the older versions of the library files. New standard development should encourage the use of the latest version of the library components so that a common language is used.
4. **Change Management.** As part of the single file schema approach, each component will be versioned separately in the XML Registry and Repository so that changes in a component can be easily identified. The schema owner will then be able to decide if the changed components are appropriate for updating the schema.
5. **Component Identification.** Since the single file schema includes only those elements that are used in the schema, the file can be used as a complete and inclusive list of components.

The one disadvantage of this approach is that the same name cannot be used for components in a single file schema. While this could be a problem especially across sector libraries, this restriction may actually promote less confusion by not allowing ambiguous component names. The TAB's experience in converting current standards to single file schemas is that this limitation can be resolved without any effect on the validation of instance documents.

Single File Schema Implementation Approach

It is recommended that PESC adopt single file schema in two phases. In Phase I, single file schemas will be generated from the existing standards by transforming those standards using an XSLT script. These single file schemas will then be stored for retrieval on the standard web site (and/or the new edu1world site). Standards development will continue using the current procedures including putting new components into the sector and core libraries. Once the standard is submitted for public comment, the single file schema will be created and made available for comment.

Phase II will use the XML Registry and Repository (R&R) instead of schema files to generate standard schemas. For existing standards, any components that are not in the

registry and repository will be added to the appropriate library (Sector and Core). For this system to work seamlessly, the definition of the root element will need to be in the repository so that the single file schema can be generated by selecting just the root element. For new standards development, the workgroup would start by creating new components in a file and then uploading all the additional components from the R&R. When the standard is available for public comment, the new components will be put in the R&R and the single file schema generated by selecting the root element component using the custom schema development feature. This step will assure that all elements are available in the R&R and that the generated schema can validate instance documents submitted by the workgroup.

Single File Schema Workgroup

While there are only a few issues that need to be resolved to accomplish Phase I, there are significant and in some cases challenging problems to resolve for Phase II. It is the recommendation of the TAB that a new workgroup be established to implement the single file schema approach. This workgroup would be responsible for planning and implementing the single file schema approach. This workgroup would work closely with FSA staff to assure that the R&R has the appropriate functionality to support the creation of single file schemas. In addition, this workgroup will be the PESC focal point for developing requirements for future enhancements of the R&R.

Conclusions

There is consensus on the TAB that the component based single file schema approach will significantly increase the reuse of components and will make it easier for the education community to adopt PESC standards.