Election Plumbing Standards: Data Format Requirements for Inter-operability, Data Publication and Election Auditing

John L. McCarthy[1], Election Auditing Project

Verified Voting Foundation

EXECUTIVE SUMMARY AND INTRODUCTION

The draft 2007 Voluntary Voting Systems Guidelines (VVSG)[2] proposed by the U.S. Election Assistance Commission (EAC)'s Technical Guidelines Development Committee (TGDC), with technical support from the National Institute of Standards and Technology (NIST) included several recommendations that voting systems hardware and software must provide detailed data in standard formats to support inter-operability, transparent reporting, and post-election audits comparing hand-eye manual counts of voter-verified records with electronic tabulation results.

This document discusses why and how the next VVSG needs to be strengthened in terms of requiring all voting systems to be able to input and output data using a *common standard format* for data import, export and exchange. Requiring a *single standard* data exchange format could also facilitate another important VVSG goal -- interoperability of election hardware and software components from different vendors.

The draft 2007 VVSG "encourages" adoption of a standard data exchange format to facilitate interoperability between different hardware components, but that is not enough. We strongly urge the EAC to strengthen that aspect of the draft 2007 VVSG by *requiring* voting systems to support input and output using the Election Markup Language (EML), the one existing consensus-based standard for election data that has been developed over the past eight years by the well-known OASIS international web standards development consortium, and that has already been widely tested and used. We also emphasize another important reason for using EML -- i.e., quick transmission of detailed, precinct-level election results from local, statewide, and multi-jurisdiction contests to state election offices, the media, and the public in order to

---

[1] John McCarthy has been a volunteer with Verified Voting Foundation since 2004. He worked as a computer scientist on database and metadata management projects at Lawrence Berkeley National Laboratory from 1980 to 2003, and still works there part time on an ISO metadata registry standards project. He was a member of international standards committees for data dictionaries (ANSI), XML schema (W3C) and metadata registries (ISO). In May, 2008, the authors submitted an earlier version of this paper to the Election Assistance Commission in response to its request for comments on the 2007 Voluntary Voting Systems Guidelines (2007 VVSG).

[2] http://www.eac.gov/vvsg/

provide well-specified, easy-to-use data for timely reporting and post-election auditing prior to certification of final election results.

WHAT DATA SHOULD BE REQUIRED FROM VOTING SYSTEMS?

Voting systems deal with many different kinds of data, including:

- Election districts and district boundaries
- Candidate Nomination, Response to Nomination and Approved Candidate Lists
- Referendum Options Nomination, Response to Nomination and Approved Options Lists
- Voter Registration information, including eligible voter lists
- Ballot information (races, contests, candidates, etc.)
- Voter Authentication
- Vote Casting and Vote Confirmation (including information about vote casting devices)
- Election counts, results, and statistics
- Audit information pertinent to various of the above categories

Ideally, all these different kinds of data should be represented in a single, well-defined representational framework that minimizes redundant (and potentially inconsistent) information, re-uses data that recur in different places, and makes the data available for a variety of purposes. Having well-defined and controlled data can help reduce opportunities for errors as well as unnecessary costs. Although we will focus mainly on data required for election auditing, many of the same points apply to data that are used for different purposes.

The current draft 2007 VVSG specifies most, but not all of the data types needed to support minimum quality vote-tabulation audits. Extracts from relevant sections of the VVSG are shown below in Exhibits 1 and 2.

EXHIBIT 1: VVSG Part 1, chapter 4.3.2-A Tabulator, summary count record

*Each tabulator **SHALL** produce a Tabulator Summary Count record including the following:*

a. Device unique identifier from the X.509 certificate;

b. Time and date of summary record;

c. The following, both in total and broken down by ballot configuration and precinct:

1. *Number of read ballots;*
2. *Number of counted ballots;*
3. *Number of rejected electronic CVRs; and*
4. *For each N-of-M (including 1-of-M) or cumulative voting contest appearing in any ballot configuration handled by the tabulator:*
   - I. *Number of counted ballots that included that contest, per the definition of K(j,r,t) in **Part 1:Table 8-2.**;*
   - II. *Vote totals for each non-write-in contest choice per the definition of T(c,j,r,t) in **Part 1:Table 8-2.**;*
   - III. *Number of write-in votes;*
   - IV. *Number of overvotes per the definition of O(j,r,t) in **Part 1:Table 8-2.**; and*
   - V. *Number of undervotes per the definition of U(j,r,t) in **Part 1:Table 8-2.**.*

In addition to the reporting context corresponding to the tabulator itself, reporting contexts corresponding to the different ballot configurations handled by that tabulator are synthesized. These contexts are quite narrow in scope as they include only the ballots of a specific configuration that were counted by a specific tabulator. The tabulator is not required to handle the complexities of reporting contexts that are outside of its scope.

EXHIBIT 2: VVSG Part 1, 4.4.2.4-E "VVPAT, cut-sheet, content requirements per electronic CVR" summarizes other types of context information required for each count record, including:

a. Polling place;
b. *Reporting context*;
c. Date of election;
d. *Ballot configuration*
e. Type of voting (e.g., provisional, early, etc.);
f. Complete summary of voter's choices;
g. For each ballot *contest*:
   1. *Contest name (e.g., "Governor");*
   2. *Any additional information needed for unambiguous interpretation of the VVPR;*
   3. *A clear indication, if the contest was undervoted; and*
   4. *A clear indication, if the choice is a write-in vote.*
h. *An unambiguous indication of whether each sheet has been accepted or rejected by the voter.*

OTHER DATA REQUIREMENTS THAT NEED TO BE ADDED TO THE VVSG

Several important points need to be added to the 2007 VVSG sections outlined above. The EMS system needs to store information in a way that enables it to determine and report at various levels of specification including: casting methods (e.g., DRE, in-precinct scanner, central count scanner), precinct identifier, batch number (i.e., for batches of absentee ballots counted in batches on precinct or central scanners), and individual device identifier. Meaningful audits require EMS reports with batch and precinct subtotals as well as machine-readable data (in EML format) that include this level of detail -- and that match the physical storage of batches of ballots that may be selected for auditing.

Data also should be recorded and kept at a level of granularity that corresponds to the lowest level audit unit. In general, the smaller the audit unit, the more cost efficient the audit. For lower-capacity voting devices, the device level is probably the best level of granularity -- as opposed to the level of individual VVPAT-rolls, which might be difficult for the machine to keep track of. High-capacity devices such as central-count optical scanners should record subtotals at the batch level as well as for precincts. These requirements should be covered by Part 1:4.2.2-A.1 of the VVSG.

Centralized voting systems should support batching and locating types of physical ballots to facilitate efficient post-election auditing. For example, if a jurisdiction is performing a precinct-level audit, it will need to locate all the ballots for precincts selected for auditing. To help facilitate auditing of Vote By Mail (VBM) ballots, which many jurisdictions scan centrally in batches without sorting into precincts, the Election Management System (EMS) or other software used to tabulate absentee ballots should provide a machine-readable report of the number of ballots by precinct for each batch (to help locate ballots for manual audits), as well as the total number of ballots and sub-totals for each election contest for each batch.

Comparing the number of ballots according to the e-poll books (or paper sign in records) and the ballots cast is crucial for post election audits to make sure no ballots have been added or subtracted. Therefore, the electronic ballot accounting information from e-polls, where applicable, should also be included.

When audits of election results take place, the audit results, including detailed description of how any discrepancies were resolved, should also be reported in a standard format.

WHY DO WE NEED STANDARD ELECTION DATA FORMATS?

One local election official has recently commented that "Establishing common [data exchange/integrated development environment] standards is among the most important things the

VVSG could accomplish."[3] And that sentiment is shared by many statisticians, election integrity advocates, and even a number of voting systems vendors.

Many states currently have heterogeneous voting systems that include a variety of different vote capture devices and software from multiple vendors. Many election contests span election administration jurisdiction boundaries -- including most federal, statewide, and most state legislative offices. Even a single local jurisdiction may use several different types of vote capture devices -- e.g., precinct-based optical-scan machines for precinct voting, touch-screen machines for some disabled voters, and central count optical-scan machines for absentee and provisional ballots. On election night and shortly thereafter, election results from different devices and local jurisdictions need to be quickly and accurately collected, aggregated, and managed for use in calculations (such as total number of votes and margin of victory) for reporting to the public as well as post-election audits prior to certification of final results. But those tasks are difficult if not impossible if election officials have to contend with data in many different formats, some of which cannot even be easily processed by other computer software.

Several voting systems currently export data in PDF print formats that humans can be read, but which cannot be easily read by other computer programs.[4] Timely post-election audits and reporting require formats that are not only easy for humans to read, but that can also be quickly and easily read into spreadsheets and other external software without transcription or hand-editing. Formats that cannot be easily exported and easily manipulated create substantial barriers to audits and other analysis of post-election data (e.g., by the media and academic researchers).

The current draft VVSG suggests several important improvements regarding data interchange and export formats, but then (surprisingly) it fails to propose requiring all election systems components to be able to input and output data using the single existing consensus-based standard for election data, namely the Election Markup Language (EML). As noted in its introductory discussion of how the proposed 2007 VVSG differs from what the EAC adopted in the 2005 VVSG, one important new requirement pertains to data interchange and export:

> "Requirements dealing with making voting device interfaces and data formats transparent and interchangeable have been added to Part 1:6.6 'Integratability and Data Export/Interchange.' Although these requirements do not mandate a specific standard data

---

[3] Comment by George Gilbert (Local Election Official) on on *Part 1, Chapter 4.3.1-A* dated *January 30, 2008* http://www.eac.gov/vvsg/comments/

[4] U.C. Berkeley Statistics Professor Philip Stark and Berkeley PhD Student Joseph Lorenzo Hall recently worked with Marin County, California, which uses a Premier GEMS election management system. They say it would have taken a substantial amount of custom programming to read the GEMS-created PDF files into a simple spreadsheet program for subsequent auditing analysis, so they opted to just transcribe the small dataset by hand.

> format, manufacturers are encouraged to use consensus-based, publicly available formats such as the OASIS Election Markup Language (EML) standard [OASIS07] or those emanating from the IEEE Voting System Electronic Data Interchange Project 1622 [P1622]. "

Furthermore, Part 1, Section 6.6 "Integratability and Data Export/Interchange" of the 2007 draft VVSG begins with the following statement of purpose and rationale:

> "The requirements in this section deal with making voting device interfaces and data formats transparent and interchangeable. The advantages of transparency and interchangeability include that systems and devices may work across different manufacturers and that data can be conveniently aggregated and analyzed across different platforms. The requirements address (a) integratability of hardware and (b) common public formats for data. The requirements in this section do not address or mandate true interoperability of interfaces and data, however they reduce the barriers to interoperability.
>
> *Integratability* deals with the physical and technical aspects of connections between systems and devices, which include hardware and firmware, protocols, etc. Basic integratability of devices is achieved through use of common, standard hardware interfaces and interface protocols such as USB. Thus, a printer port must not be proprietary; it must use a common hardware interface and interface protocol, with the goal being that printers of similar type should be interchangeable.
>
> Systems and devices that are integratable are designed such that components of systems may be compatible or can be made compatible with each other through some moderate amount of effort, for example, by writing "glue code." For example, an audit device may be designed to work with a DRE, but it may require adaptations to protocols for signaling or data exchange. Adapting the audit interface to the DRE may require some amount of software modification but should still be within reasonable bounds.
>
> The barriers to interoperability are further reduced if all systems support the same commonly agreed upon, publicly-available data format for ballot definition, records and reports. The advantages to using common data formats include:
>
> - Common formats for specifying election programming data such as ballot definition files promotes greater accuracy and reduces duplication;
> - Common exported data formats can assist in aggregating results and conducting analyses and audits across among manufacturer systems; and
> - Common formats for use in data reports can be mapped as necessary to locality-specific reports as opposed to requiring the device to export the report in the locality-specific format."

Subsequent parts of section 6.6 require "a non-restrictive, publicly-available format" for Election Management System (EMS) software that deals with election programming and report data, and for export of Cast Vote Records (CVRs) from Direct Recording Electronic devices (DREs) and optical scanners.

COMMON DATA FORMATS ALSO ARE NEEDED FOR BALLOT DEFINITION

Prior to each election, local election officials (or vendors under their direction) create ballot definition files that specify the content and format of ballots that voters use to cast their ballots -- either paper ballots, or ballot displays on some kind of touchscreen device, or via other mechanisms (e.g., telephone). These ballot definition files are currently written in vendor-specific computer languages, often using some kind of graphical user interface software.

As noted above, Part 1, Section 6.6 of the draft 2007 VVSG observes that "Common formats for specifying election programming data such as ballot definition files promotes greater accuracy and reduces duplication" as well as reducing barriers to interoperability. Yet the section that deals specifically with ballot definition (7.1), does not mention the need for a common ballot definition language. Section **7.1-B** "EMS, ballot definition" says simply that "The EMS ***SHALL*** provide for the logical definition of the ballot, including the definition of the number of allowable votes for each contest.

Requiring EML (which is a dialect and extension of XML) for ballot definition could go a long way toward improving ballot design in the United States, where poor ballot design has been the source of a number of notorious voting problems (e.g., the "butterfly ballot" in 2000). One of the primary strengths of XML is to separate content from layout in formatting documents and text. In the case of ballots, a single underlying XML document could be used to specify and generate paper ballots, sample ballots, touch-screen displays, etc. simply by using different XML Style Sheets. Because there are a large number of software programs that can render XML using style sheets, and because there are many people familiar with using different variants of XML, election officials would have a much broader base of people and resources to help with ballot preparation. And having a single underlying standard for ballot components would facilitate adoption of best practices for ballot design, definition, and implementation.

HOW ARE DATA STANDARDS SECTIONS OF DRAFT VVSG INADEQUATE?

As discussed in more detail below, we feel that it is not sufficient for the EAC to just "encourage" vendors "to use consensus-based, publicly available formats." Experience with other data standards suggests that simple "encouragement" is unlikely to succeed. Part 1, Section **6.6 B.6** states that "The voting system manufacturer ***SHOULD*** use a common format for export and interchange of data and reports across its major device categories," but use of the word "should" rather than "shall" means that use of an open, standard data format is not required, and a voting systems vendor can use multiple formats across different types of hardware.

Rather than requiring support for a single common standard data format, the current 2007 draft VVSG uses different language in different places with regard to data formats. Part 1:4.3.1-A requires that audit records be available in a “fully specified, public format”. However, “fully specified” and “public” do not necessary correspond to “open,” and they are subject to differing interpretations. Part 1:4.4.1-H is more specific about the requirements for an open format being “non-proprietary” and “requiring no special knowledge of confidential or proprietary or trade secret information”. But none of these wordings include the need to provide machine readable

and machine-processable output to support auditing. For example, the discussion for 4.3.2-A specifies that the “[tabulator] record must be output in a human-readable format” but it says nothing about machine-readability or -processability.

A number of technical experts have already made comments about the shortcomings of the proposed data requirements using the EAC's web-based tool for submitting section-specific comments.[5] Several have noted that the current proposal is "Very Subjective, that "Publicly-available format" and "non-restrictive" are not well defined(and perhaps cannot be), and even that "As there is currently no "common consensus based format" it is inappropriate to set a requirement for one" -- which is not even true (because EML is clearly such a standard)..

"[D]escriptions of elements, attributes, constraints, extensions, syntax and semantics of the format, and definitions for data fields and schemas," as called for in section **6.6 B.4** are challenging to create even in the best circumstances with plenty of resources and talented professionals. Most vendors already have such metadata that specify their own interchange formats, but those have not been easy for others to use. Adoption of a single, recognized metadata specification standard such as EML would make much more sense for all election data stake-holders.

It is surprising that the TGDC 's draft 2007 VVSG did not at least propose requiring use of XML (eXtensible Markup Language) for data exchange and export, since XML has become the ubiquitous open, publicly-available, worldwide standard for data exchange and export of all kinds of data.[6] XML is self-describing -- that is, XML tags (as shown in Figure 1 of Appendix A below) provide human-readable labels of each type of data item and data structure, using the simple, yet elegant and extensible XML syntax that is now widely used by a variety of existing hardware and software all over the world. XML is a natural choice for a standard, non-restrictive, publicly-available format for election data. Even better, there is already a dialect of XML that has been developed over the past seven years specifically for election data, namely the Election Markup Language (EML), which is described in more detail below.

Unfortunately, section 6.6-B.6 in Part 1 of the current draft VVSG is a recipe for continuing delay and frustration. It gives vendors little incentive to use a truly non-restrictive, publicly-available format, and it is not likely to give election officials, the media, and the public the kind of well-specified data formatted in a standard way that is necessary to do timely, detailed reporting and auditing. In fact, the vague proposed "standards" for data would continue to make integrated statewide reporting and inter-operability difficult if not impossible for states whose jurisdictions have multiple election systems vendors.

---

[5] http://www.eac.gov/vvsg/comments/ select chapter 6 and browse the comments

[6] http://www.w3.org/XML/

WHY SHOULD THE EAC MANDATE OASIS EML AS PART OF THE VVSG?

EML (Election Markup Language) is a dialect of XML. XML (eXtensible Markup Language) is the *lingua franca* of the World Wide Web, and it is the basis of thousands of specialized "dialects" such as EML. EML is a comprehensive, open, extensible, non-restrictive international standard comprised of data and message definitions described in a coordinated, modular set of XML schemas. That is, EML is a specific set of XML data structures, elements, and attributes specified by a series of modular XML schemas.

EML has been developed by the OASIS Election and Voter Services Technical Committee over a period of eight years since 2001.[7] This technical committee has included representatives from major voting systems vendors and government officials, as well as computer scientists and metadata experts.[8] It has been chaired throughout its existence by a representative of the British Government.[9] In January, 2008, OASIS members voted to approve the fifth major release of EML (5.0) to replace EML 4.0 as an official OASIS standard, and current work on the new EML 6.0 includes many new additions, primarily for United States election systems.

OASIS itself (Organization for the Advancement of Structured Information Standards) is a not-for-profit consortium that drives development, convergence and adoption of open standards for the global information society. OASIS produces more Web services standards than any other organization, along with standards for security, e-business, and standardization efforts in the public sector and for application-specific markets. Founded in 1993, OASIS has more than 5,000 participants representing over 600 organizations and individual members in 100 countries.

OASIS is currently in the process of submitting EML to the International Standards Organization (ISO) for it to become a full ISO standard. It is expected that the Joint Technical Committee 1 will assume responsibility for EML and manage future updates in conjunction with OASIS.

---

[7] For an excellent brief (21 page) overview of EML, including references to case studies of where EML has been used to date, see the OASIS White Paper "The Case for using Election Markup Language (EML) http://www.oasis-open.org/committees/download.php/26747/The%20Case%20for%20EML%20v2.pdf

[8] See http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=election

[9] John Borras <johnaborras@yahoo.co.uk>

In 2004, the Council of Europe recommended using EML to its 46 member states,[10] and various European countries and election systems vendors are gradually adopting EML. In October, 2007, five vendors (including Oracle, IBM, and ES&S) conducted a successful interoperability demonstration of voter registration, different types of voting "channels" (equipment), and counting/results in London that showed how EML and XML can facilitate inter-operability between different parts of the election process as well as among different vendors.[11] In February, 2008, after eight months of development and testing, the California Secretary of State's Office successfully used EML to report statewide election results from the February 5 Presidential Primary election in that state.[12]

EML 6.0 currently includes specifications (in separate modular schemas) for:

- Election districts and district boundaries
- Candidate Nomination, Response to Nomination and Approved Candidate Lists
- Referendum Options Nomination, Response to Nomination and Approved Options Lists
- Voter Registration information, including eligible voter lists
- Various communications between voters and election officials, such as polling information, election notices, etc.
- Ballot information (races, contests, candidates, etc.)
- Voter Authentication
- Vote Casting and Vote Confirmation
- Election counts, results, and statistics
- Audit information pertinent to some of the other defined data and interfaces

EML is flexible enough to be used for elections and referendums that are primarily paper-based or that are fully e-enabled, and the OASIS Technical Committee plans to continue extending it as needed (and issuing subsequent versions of EML) in the coming years.[13]

In addition, OASIS EML V6.0 includes templates and a dictionary system that make tailoring each component for localization needs dramatically faster and easier

Figure 1 in Appendix A shows an example extract from the beginning of a test EML file for the California 2006 Governor's election (using EML schema 510 for vote counts), while Figure 2 shows a web display table generated from the EML file using the stylesheet shown in Figure 3.[14]

---

[10] Election Markup Language (EML) Recommended to Member States by Council of Europe Cover Pages, 10 Nov 2004 http://xml.coverpages.org/ni2004-11-10-a.html

[11] http://www.oasis-open.org/committees/download.php/25992/EML%20Interop%20Demo%20Report.pdf

[12] http://www.sos.ca.gov/media/

[13] http://www.oasis-open.org/specs/index.php#eml5.0

Although EML includes a number of components that may not be necessary for current American election systems, it is quite modular, and it includes a full subset of features that can be used immediately for most, if not all, of the components called for in the current draft 2007 VVSG. EML is also extensible, so it can be easily adapted to include whatever additional kinds of data may be needed that it does not already include.

Having election data publicly available in an EML format would make it relatively easy for anyone with modest computer programming skills to write free, open-source software for putting local and state data up on web pages almost immediately on election night, which could be available for audit sample selection as well as providing RSS feeds to media and anyone else who wishes to use it for further analysis. This has been already demonstrated by the state of California's use of EML 530 transactions for news media in 2007 and 2008 elections.

POSSIBLE OBJECTIONS TO REQUIRING EML IN THE 2007 VVSG

Some possible objections to requiring EML as part of the 2007 VVSG are that it is too new, that other competing standards for election data may be better suited for American elections, that EML includes a number of constructs that are not currently necessary for most American elections while lacking others that are necessary, that has not been sufficiently tested in real elections, and that it would be difficult and costly for vendors to implement. This section considers each of these issues in turn.

Is EML Too New? As noted above, the OASIS Election and Voter Services Technical Committee and others have been developing and testing EML over a period of eight years since 2001. There have already been five major releases of EML to date, and it has already been used in production for several real elections as well as distributing election results.

Although EML is a relatively new standard, it has been tested and used for a number of different purposes in recent years. Over the past few years the United Kingdom has conducted a number of e-voting pilot studies with favorable results.[15] EML was used in local elections in Flanders (Belgium) in 2006 and 2007.[16] The Australian Electoral Commission (AEC) has developed an election results publishing system based on EML. The AEC's "Media Feed" (as it is called) provided near real-time results to a wide variety of media, from individual bloggers to major media organizations during the 2007 Australian Federal Election. Given the success of its Media

---

[14] example data from http://www.sos.ca.gov/media/

[15] http://www.justice.gov.uk/guidance/may2006electionpilots.htm, www.dca.gov.uk/consult/core/cp2905.htm , http://www.justice.gov.uk/guidance/may2007electoralmodernisation.htm,

[16] www.oasis-open.org/apps/org/workgroup/election/download.php/20745/LV2006_Local%20Elections%20Flanders%202006-v0%2031_V01%5B1%5D.00.pdf

Feed, the AEC is likely to draw upon EML in forthcoming projects to drive election systems and deliver election results.[17]

Competing Standards? When NIST staff and the Technical Guidelines Development Committee (TGDC) began to discuss the new VVSG early in 2007, EML 4.0 was lacking some important features, and it appeared there might be another competing standard, the IEEE Voting Systems Electronic Data Interchange Project 1622. But several months after the TGDC issued its new draft VVSG in September, 2007, IEEE Project 1622 was temporarily deactivated because the technical committee "failed to achieve balance" (no one interest group can constitute fifty percent or more on IEEE technical committees).[18] Before it was suspended, Project 1622 had a substantial overlapping membership with the OASIS committee and was working on an XML-based model that closely resembled EML. Moreover, as noted above, OASIS has now released a new version of EML (5.0) that addresses many of the shortcomings of EML 4.0, and is now completing work on EML 6.0, which has focused particularly on US style elections and needs.

Why Not Just Require Comma-Separated Values? Some people have suggested that requiring comma-separated values might be sufficient for data export and exchange, but that approach is insufficient in several respects. First, as the VVSG discusses in some detail, data require clear definitions for what each and every data field means, and comma-separated values do not provide a natural way to automatically include such metadata. Second, as Exhibits 1 and 2 above illustrate, election data has a number of hierarchical components that would require a complex set of inter-related tables, with definitions for each column in each table. Third, it would be very costly and error-prone to write separate scripts for each state and jurisdiction to try to interpret comma-separated values from different hardware and databases.

Because it is defined using a set of inter-related, modular XML schemas, EML already has clear and well-specified definitions for each and every data structure, data element, and attribute. If voting systems vendors are required to support EML for input and output of all data, then they alone will be responsible for making sure which data in their systems gets put into which EML data elements. In fact, a number of vendors already use EML or some similar kind of XML data structures for a number of their products, so they are already familiar with EML and XML. Requiring EML would be a relatively small and logical next step for the VVSG to mandate.

Unnecessary Features? Although EML has more components than needed for most current American elections, the modular design of EML makes it relatively easy to use just a subset of EML sub-schemas necessary for a particular function, type of hardware, or type of software. Each election administration jurisdiction can tailor EML to its own specifc needs using

---

[17] http://www.aec.gov.au/media/mediafeed

[18] private email from Bill Ash, Senior Program Manager, IEEE Standards Activities Dept (3/18/2008)

Schematron[19] or other similar softare. The UK Government did such an exercise to support its e-voting pilot studies.[20]

Missing Features? Since EML is based on a modular set of XML schemas, it also can be readily extended by adding new elements, attributes, and schemas when necessary. For example, most information needed for post-election vote tabulation auditing is contained in the EML 510 and 520 schemas, while the new EML 530 schema was developed in response to needs of the California Secretary of State's Office in 2007 for its experimental use of EML for reporting statewide results from the February 5, 2008 primary election in California. EML 510, 520 and 530 all now share a common information model core structure that fully supports US style election reporting.

Implementation Costs. In terms of implementation cost, it probably would be less time-consuming and expensive for vendors to implement the necessary EML modules that have been carefully designed, developed and tested, rather than each trying to develop their own formats and specifications. In fact, at least three major American vendors already use EML or XML in some significant way. ES&S already uses EML for some of its products. Hart-Intercivic uses a dialect of XML (called EDX[21]). Premier uses XML for its Voter Registration product. ES&S's Peter Zelechoski, who is an active member of the OASIS Election and Voter Services Technical Committee and chaired the IEEE Voting Systems Electronic Data Interchange Project 1622 until recently, stated that ES&S would in fact *prefer* that the EAC *require* EML for data export and interchange, rather than the proposed much more vague requirement for each vendor to create it's own idiosyncratic data descriptions, which would very likely create a lot of needless frustration, argument, and even litigation.

BENEFITS OF REQUIRING EMLFOR VVSG DATA STANDARDS

The preceding sections have discussed a number of operational reasons for and questions about requiring EML as a data exchange standard within the VVSG, including support for ballot definition, reporting, and post-election auditing. But it may be helpful to also consider potential benefits in terms of somewhat more abstract aspects of election administration, including Interoperability, Accuracy, Ease of Administration, Accessibility, Security, Testability, and Cost.

Interoperability - As noted above, Part 1, section 6.6 of the proposed draft 2007 VVSG discusses how data exchange standards can help reduce barriers to inter-operability, which can in turn make it easier to aggregate and analyze data across different hardware and software

---

[19] The Schematron Assertion Language 1.5 (ISO/IEC 19757-3:2006) see www.ascc.net/xml/schematron/

[20] The UK Localisation of EML see http://www.legsb.gov.uk/blueprints/item.php?id=173

[21] http://www.hartic.com/pages/157

components of a voting system. Interoperability also gives election officials more flexibility to "mix and match" components from different vendors in order to conduct cost-effective, accurate, and secure elections.

Accuracy – Standard data interfaces between election system components could improve accuracy in several different ways, particularly for ballot definition, reporting, and auditing. One ballot file could be used for ballot printers (for normal, absentee and provisional ballots), optical scan machines, ballot marking devices, touch screen devices, and tabulation software. In the current non-standard environment, every time a different ballot definition file needs to be programmed, another potential for error is introduced. Also, if there were only one specified ballot file for each contest, it would be have more scrutiny and less chance for error. It would also create more incentives to create and use ballot definition templates that could incorporate best practices for ballot layout, candidate rotation, etc. Likewise, requiring EML for data will make it easier to develop templates and open source software for reporting and auditing.

Ease of administration – Many headaches of administering elections concern ballot definition files. Standard ballot definition files would result in fewer problems and would ease the administration of elections. Moreover, election officials would be more able to mix and match systems for final tabulation and would find it easier to introduce equipment upgrades and changes.

Testability – If each election system component (hardware and software) was required to use EML for input and output, it would become much easier to test individual components independently, which could in turn help reduce testing time and costs.

Accessibility – Standardized and public data interfaces could help encourage more innovations for specialty equipment because specialty equipment or accessibility add-ons could be more easily tested independent of other election system components. Also, standardized and public data standards could help ensure that ballots for voters with disabilities would have sufficient scrutiny to ensure their accuracy.

Security – Requiring each election system component to use EML for input and output would also enable systems to be integrated and deployed with components from different vendors. Not relying on one vendor for every component in the system reduces the possibility of a security attack from within a particular company..

Cost – Last, but not least, requiring standardized public formats for election data could also help bring down the cost of elections. Election officials would have more flexibility to shop around for the best value for each component of their voting systems. There would be lower barriers to entry if new entrants into the market could introduce one component of a system instead of having to develop an entire integrated system. With standardized formats, especially for ballot definition files, election administrators would no longer have to rely primarily on system vendors and contractors to program various machines for each election. For voting system vendors, the cost of creating data interfaces for EML will be less expensive than each vendor designing, developing, and documenting their own idiosyncratic "standard" formats -- because EML data

structures are already well-specified and documented, having been carefully designed and developed by a number of experts (including vendor representatives) and tested in a number of different voting jurisdictions over several years.

RECOMMENDED CHANGES TO REQUIRE EML FOR 2007 VVSG

The current draft 2007 VVSG requirements, recommendations, and "encouragement" in various sections of the VVSG to use "a *non-restrictive, publicly-available format"* is not specific enough nor standard enough to accomplish the goals of software independence, inter-operability, ease of use, and cost minimization enunciated elsewhere in the VVSG. In order to achieve those goals and others discussed above, the VVSG should require support for the OASIS Election Markup Language (EML version 5.0 or higher) for all data input, output and data exchange between system components throughout the entire elections process, from candidate filing and ballot definition through election reporting and auditing,

A number of specific sections in the current draft 2007 VVSG need to be revised to REQUIRE that the OASIS Election Markup Language (EML version 5.0 or higher) must be supported for all data input, output and data exchange between system components. Sections in need of such revision include 6.6, 7.1, 4.2 and 4.3. For a detailed list of specific comments and suggested changes for particular sections of the current draft 2007 VVSG, please see https://vvf.jot.com/PublicVVSGcomments -- which we have already entered in the EAC's excellent web-based on-line system for making comments on the VVSG.[22]

CONCLUSIONS

Data standards should not be considered an add-on or extra cost. They should be an essential element for all voting systems. A number of sections in the current draft 2007 VVSG point out the importance and advantages of using a consensus-based data standard for various hardware and software components of election systems. Fortunately, the Election Markup Language fits all the criteria that the VVSG outlines for such a data standard. It has been carefully developed over a number of years by an international OASIS Technical Committee that has included representatives from vendors as well as election officials and election data specialists.

Several major voting system vendors already support EML or XML dialects that resemble EML for at least some of their products, including Election Systems and Services (ES&S), Hart Intercivic, and Premier (formerly Diebold). Representatives from all of these three major American voting systems vendors have said that they would welcome having the EAC require EML for data exchange and export, rather than the more generic and vague requirements proposed in the current draft 2007 VVSG.

[22] http://www.eac.gov/vvsg/comments/

With adoption of key additions proposed by the IEEE P 1622 effort in 2007, EML 5.0 (which OASIS members approved in January, 2008) now has a sufficiently mature, extensive, and extensible feature set that the EAC has no reason to delay in requiring its adoption as an integral part of the 2007 VVSG. The EAC should REQUIRE all voting systems to be able to read and write all data in standard EML data formats (version 5.0 or higher) in order to expedite both voting systems interoperability and easy export and aggregation of all election results for timely reporting and rigorous, cost-effective election auditing.

ACKNOWLEDGEMENTS

Thanks to the statisticians and election experts from the Election Auditing Group that has been meeting since January 2007 and others (including several vendor representatives) for many helpful comments and suggestions that helped improve this paper. Special individual thanks for encouragement, helpful comments and suggestions from David Dill, Bob Kibrick, David Webber, John Borras, Peter Zelechoski, Richard Johnson, and Joe Hall.

APPENDIX A: Example EML and related files

Figure 1: First Part of EML data file

```
<?xml-stylesheet type="text/xsl" href="cal510_style.xsl"?>
<EML Id="510" SchemaVersion="5.0">
   <TransactionId>69</TransactionId>
   <MessageLanguage>en-US</MessageLanguage>
   <IssueDate>2006-12-14T20:06:12.003</IssueDate>
   <Count>
      <EventIdentifier Id="2006_California_General_Election"/>
      <Election>
         <ElectionIdentifier Id="20061107"/>
         <Contests>
            <Contest>
               <ContestIdentifier Id="020000000000" DisplayOrder="2">
                  <ContestName>Governor</ContestName>
               </ContestIdentifier>
               <TotalVotes>
                  <CountMetric Id="PR" Type="Precincts Reporting">25090</CountMetric>
                  <CountMetric Id="TP" Type="Total Precincts">25090</CountMetric>
                  <Selection>
                     <Candidate>
                        <CandidateIdentifier Id="122" DisplayOrder="1">
                           <CandidateName>Phil Angelides</CandidateName>
                        </CandidateIdentifier>
                        <Affiliation>
                           <AffiliationIdentifier Id="Party">
                              <RegisteredName/>
                           </AffiliationIdentifier>
                           <Type>Democratic</Type>
                        </Affiliation>
                     </Candidate>
                     <ValidVotes>3376732</ValidVotes>
                     <CountMetric Id="PVP" Type="Percent Votes in Party">100.0</CountMetric>
                     <CountMetric Id="PVR" Type="Percent Votes in Race">39.0</CountMetric>
                  </Selection>
                  <Selection>
                     <Candidate>
                        .....
                  </Selection>
```

Figure 2: Part of Web Table Generated from EML file in Figure 1

- Election Results

2006_California_General_Election 20061107

Data Timestamp: 2006-12-14T20:06:12.003

Ballot	Metric	Total
Governor	**ContestID = 020000000000**	
STATEWIDE		
	Precincts Reporting	**25090**
	Total Precincts	**25090**
Phil Angelides [Party=Democratic]		**3376732**
	Percent Votes in Party	100.0
	Percent Votes in Race	39.0
Arnold Schwarzenegger [Incumbent] [Party=Republican]		**4850157**
	Percent Votes in Party	100.0
	Percent Votes in Race	55.9
Edward Noonan [Party=American Independent]		**61901**
	Percent Votes in Party	100.0
	Percent Votes in Race	0.7
Peter Camejo [Party=Green]		**205995**
	Percent Votes in Party	100.0
	Percent Votes in Race	2.3
Art Olivier [Party=Libertarian]		**114329**
	Percent Votes in Party	100.0
	Percent Votes in Race	1.3
Janice Jordan [Party=Peace and Freedom]		**69934**
	Percent Votes in Party	100.0
	Percent Votes in Race	0.8
Alameda County		
	Precincts Reporting	**1219**
	Total Precincts	**1219**
Phil Angelides [Party=Democratic]		**229217**
	Percent Votes in Party	100.0
	Percent Votes in Race	56.6

Figure 3: Part of XSL Stylesheet Used to Produce Figure 2 From EML Data in Figure 1

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
- <xsl:stylesheet version="1.0" xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
- <xsl:template match="/">
- <html>
- <head>
  <style type="text/css">th {FONT-FAMILY:Arial; font-size:10pt} td {FONT-FAMILY:Arial; font-
size:10pt}</style>
  </head>
- <body>
- <xsl:element name="img">
  <xsl:attribute name="src">vote-stats.gif</xsl:attribute>
  <xsl:attribute name="border">0</xsl:attribute>
  <xsl:attribute name="alt">Election Results</xsl:attribute>
  </xsl:element>
  - Election Results
- <center>
- <h2>
  <xsl:value-of select="//Count/EventIdentifier/@Id" />
  </h2>
- <h2>
  <xsl:value-of select="//Count/Election/ElectionIdentifier/@Id" />
  </h2>
- <h3>
  Data Timestamp:
  <xsl:value-of select="//IssueDate" />
  </h3>
  </center>
- <center>
- <table style="border-collapse: collapse" border="1" width="700" cellpadding="2" cellspacing="0"
bordercolor="#666666">
- <tr bgcolor="#9acd32">
  <th align="left" width="50%">Ballot</th>
  <th align="left" width="30%">Metric</th>
  <th align="left" width="20%">Total</th>
  </tr>
- <xsl:for-each select="//Election/Contests/Contest">
- <tr bgcolor="#dddddd">
- <th align="left" width="50%">
- <font size="4" color="#cc3333">
  <xsl:value-of select="ContestIdentifier/ContestName" />
  </font>
  </th>
```