

# ***Data Normalization in Electronic Voting Systems: A County Perspective***

*Paul Lux, CERA*

*Supervisor of Elections*

*Okaloosa County, Florida*



## ***Data Elements***

The common elements shared by election systems, specifically at the tabulation server level where the database for each individual election is laid out, may vary in structure and organization, but share a commonality of function. These are: contest identification numbers; candidate/choice identification numbers within each contest; party affiliation codes for candidates in partisan races; counters or counter groups—used or not—for each candidate/choice, over vote (if allowed), blank vote, and under vote; ballot headers common to all ballots; ballot headers specific to certain ballots; individual contest headers; and ballot footers common to all ballots.

Variations of these exist, obviously, and depending on the jurisdiction some of them may not be required and thus go unused (straight-ticket voting for example). For those who write the programs, however, all of these items must still be taken into account and be part of the design if one hopes to sell the product across many varied jurisdictions. Thus, even where certain features are not desired or legal to use, they still exist in the program and are switched off or simply go unused.

Of course each of the four major vendors of election tabulation software (identified by this author as ES&S, Hart InterCivic, Premier, and Sequoia) has their own individual sizes and data formats for each of these fields. Some limit certain fields to numeric characters, while others may allow alpha characters or even free-form fields. When this happens, sharing such important information at any central level becomes problematic.

## ***First Steps in Florida***

In 2007 the Florida Division of Elections sought to bring some form of standardization to some of these data elements to better facilitate the speed and accuracy of election night reporting. Rather than write file translation code for the results export of each of our 67 counties, they provided race ID codes for all races that were required to be reported to the State (state legislative seats and federal races).

Florida, with its own robust certification process, allows each individual county to choose the elections equipment it feels best serve the needs of its voters from the list of approved vendors. Currently, three systems are certified and in use in Florida: ES&S, Premier, and Sequoia. At the first meeting where the standardization of election codes was discussed, Dominion was also in attendance as they were seeking certification at that time.

Of the four vendors in attendance, only Dominion allowed the use of \*.eml (election markup language) to export results. The other three stated that they were “years” away from working \*.xml into their exports, let alone \*.eml. With nothing better than a standard \*.csv or \*.txt file in common, the Division of Elections settled on a four-digit numeric code (a six-digit was preferred, but ES&S had a size limitation issue at that time) that allowed for uniform representation of races, party affiliations (for use in primaries) and district numbers (where applicable). A few counties, Okaloosa among them, volunteered to use the new code system as a test run in January’s Presidential Preference Primary.

The actual mechanics of it were simple: where the tabulation server database software would automatically number entered races in a standard pattern (10, 20, 30, etc.) a county would simply substitute the specified code for each particular race. This would preclude the State from having to map 67 individual county files with different race numbers. The codes were structured such that the order of contests on a ballot as mandated by statute would be preserved.

The test was successful, and a new set of file specifications was developed for statewide use in the August Primary and the November General elections. Having the new file specifications in place made the accumulation of statewide election results a more streamlined process. Each county uploaded the same results export file as normal, but this time the State could marry up results data for contests that spanned multiple counties with little effort.

### ***Problems with Other Voting Systems***

The potential issues I see moving forward with the normalization of election tabulation data have to do with two very specific types of election systems: non-electronic voting systems using hand-counted paper ballots and other non-traditional voting systems such as vote-by-phone or voting over the internet.

The problems associated with non-electronic systems are obvious: no electronic data means no electronic file to upload or share. Of course an electronic file can be created (spreadsheet, etc.) using any field codes deemed important; but the data will, by its very nature, be hand-entered in some form or fashion which will always call its veracity into question.

As currently configured, a feature to import other non-traditional voting systems' data into tabulation servers is nonexistent. In part, this is due to voting system vendors hesitation to share their data structures with potential rivals who might later expand into their market. This leads, again, to a manual solution of getting this data into the tabulation system for reporting of results.

### ***Data Access***

The biggest question that will probably arise from this is who should have access to this new and improved form of data? Obviously there is a need for it at the level where multi-jurisdictional contests are tabulated. But what about other potential end users?

If data across the various election systems platforms becomes standardized, it would allow the media to report contest results more rapidly. The problem with allowing the media access to such files would be the source of the file—i.e. would they get it from local jurisdictions directly or collect them from the state. If only the state government can report unofficial or official results, getting these files directly from the source jurisdictions at the same time as the state central reporting office might lead to controversy if results that differ from the state or local jurisdiction are posted by the media. Arguably this might also be helpful as a double-check on the published "official" results, making error detection easier. It has happened in recent elections that results reported on local jurisdictions' websites have not matched the "official" totals. There would also be a technological learning curve on the part of the

media—a media used to hand-entering data from faxes and e-mails—to make sure the file’s contents were accurately reflected.

Right on the heels of the media for these files will be the political parties. Generally speaking, political parties are mainly interested in what benefits them. Florida is one of the most open states when it comes to public records and open meetings laws. To date this jurisdiction has never been asked for the actual results export file. Parties, to date, have been content to settle for paper copies of the data faxed or e-mailed on election night. Having the source files, again, would allow them to do their own computation of any contests’ outcome; but any discrepancies could ultimately lead to politically motivated accusations of data tampering.

### ***File Security***

One of the issues that will have to be addressed if normalized data files are to be made widely available is the security of the file itself. Too many groups with too-free access to the file could easily result in disputes as to who has the “true” copy of the file. Given the increasing regularity with which election officials have come to be distrusted, some mechanism to authenticate the file should be part of any such system.

The discussion of the merits of digitally signing the file, encrypting it or hashing it will be left to those whose background is better suited to the subject. As no such safeguards are currently employed, the addition of such measures would certainly be welcome sooner than later.

### ***Audits and Safeguards***

As such a data file might relate to audits there is plenty of room for speculation about how such data could improve audits. Some jurisdictions already scan and post each and every ballot as a \*.pdf file that can be downloaded from a website, reviewed and compared to published totals. Making this file available to the general public or to other interested groups who wish to use the data to study a particular election would go far in advancing the transparency of audits and of the conduct of the election itself. Such transparency would do a lot to repair the trust in the electoral process that has been steadily eroded on a national level since 2000.

What should be avoided at all costs, however, is the release of any data files that contain precinct-level data before the official audit is completed. In this jurisdiction the Statement of Votes Cast that details the polling, absentee, and early voting results by precinct is withheld until the statutorily mandated audit is completed so as to preclude any prejudicing of the Audit Board.

### ***Conclusion***

It cannot be denied that there would be a tremendous benefit in normalizing data file structures in electronic voting systems across the different platforms. Though issues with proprietary software and file security would have to be addressed, along with clear guidelines for how and by whom such files should be used; better accuracy in election night reporting along with increased transparency in both

the conduct of elections and election audits could be realized. The potential benefit to many diverse interests is too great to not pursue.