




Cast Vote Records – Harvie Branscomb - updated 9/20/2016

What functions will a Cast Vote Record (CVR) serve if implemented under best conditions?


- 1) Serve as an unambiguous machine and human readable record of an interpretation of the contents of a voter or machine-marked paper ballot and also voter intent ected by a DRE.
 - As machine interpreted, and alternatively
 - as human interpreted or edited by human intervention.
- 2) Also may serve as a record of voter intent both
 - as marked, i.e. prior to application of rules for election interpretation/tabulation, or
 - er application of those rules, or preferably, both. May also be:
- 3) compared visually or by machine preferably with the paper ballot itself or with a copy of the paper ballot to verify interpretation quality;
- 4) aggregated by humans or by independent machine to produce an election sub-tally; and
- 5) used to document the provenance of any interpretation of the ballot. Each CVR may:
- 6) serve as a means to detect flaws in the interpretation of voter intent;
- 7) serve as a means to characterize and quantify characteristics of voter marks;
- 8) provide efficient access to ballots subject to difficult interpretation;
- 9) allow reconciliation of vote counts to ballot count for arbitrary subsets of the election;
- 10) allow reinterpretation of an election if the voter intent rules change by court order;
- 11) assist in selecting ballots to be added to random selections for auditing; and
- 12) mit extension to include merit information about the voter marks such as mark density within target or other vendor specific metrics.

The Cast Vote Record can be best be represented as an election specific XML document containing various characteristics of the election plus separately a very efficiently encoded CSV table containing for each row or newline delimited line a set of ballot-specific fields, including for each contest two contest-specific fields, and for each contest choice a choice-specific field, each field delimited by comma. A mechanism for an optional extension of one extra field per choice/target is provided for use in detailing the merit of the detection of the mark. With this additional data, those with access can efficiently sort the Cast Vote Records according to metrics relevant to the merit of the detection of intent.

I will restrict further discussion here to the CSV formatted table(s).

The **ballot-specific fields** at the left hand portion of the CVR line contain information about the ballot including:

(values are encoded in human/machine readable UTF-8)

- The number of expected fields in the CVR – used to double check the accuracy of parsing.
- The number of fields in the ballot-specific portion of the CVR (assumed to be jurisdiction specific)
- tatewide election-scope serialized number for each paper ballot (and a corresponding but separable serial number for other non-ballot paper such as header cards) in a range assigned by

the state so that all CVRs statewide are individually enumerated and can be counted by contest and individually addressed for purposes of creating random samples for auditing.

- Election identifier (provided by state specification to avoid confusion with other elections – includes a human readable jurisdiction identifier)



Batch identifier (according to specs provided by the jurisdiction)

- Box identifier (according to specs provided by the jurisdiction)
- Sequential position of ballot in the batch (for aid in locating the ballot)
- Unique identifier as printed on the ballot at scan time or previously (if any)
- Ballot style identifier (for ease in sorting CVRs in order to easily aggregate contests)

A **per-contest bitwise field** is used for representing status and provenance of interpretation for each contest on each ballot that might for example be encoded within 64 bits as follows:

- B0-B3: bit flag incremented by one for each time CVR is modified (max 16 times)
- B4: current interpretation is an overvote condition
- B5: interpretation is incomplete (pending further resolution)
- B6-B7: 00 – undetermined; 01 - paper ballot; 02 – DRE; 03 – special
- B8- B15: may be used to characterize the ballot origin such as voting method, precinct, split, batch, etc.- jurisdiction specific and subject to ballot anonymity concerns
- B16-B31 date code and code indicating voting system or team performing the first interpretation
- B32-B47: date code and code indicating voting system or team performing the second interpretation (if any)
- B48-B63: date code and code indicating voting system or team performing the third interpretation (if any). [This structure allows for max three separate interpretations]

This column containing bitwise flags is not able to be summed either horizontally or vertically but it can be used to sort rows of the CVR table. [It may be desirable to have encoded here the number of contest choices as a cross check when parsing the table- but this can also be stored in the XML table containing the election-specific data.]

A second per-contest field is simple numerical value intended to be summed both horizontally and vertically and contains the number of additional votes that could have been tabulated in the contest if the acceptable number of choices were fully voted but are not. This number may be more than one in multi-winner contests and max is the number of winners. This contains the number of detected undervotes (if any) or the number of possible votes invalidated by overvoting. When summing within one contest a row containing this value plus all columns for choices within the corresponding contest, the number that results should match the number of vote-for choices on the contest. This can be used as a cross check of the consistency of the data if this value is taken directly from the interpretation of the marks on the ballot and not just a value derived from the detected votes.

How might Cast Vote Records display ballot contents both prior to and post application of vote interpretation rules?

Choice-specific fields follow the two contest-specific fields

Each significant mark is reflected by a number in the cast vote record- those that are to be aggregated as a vote are represented by "1" and those that are not to be aggregated as a vote are represented as a "0". Insignificant marks and no mark at all and no possibility of a mark (contest not on ballot) are all represented by blank. The CVR contains a column for each choice contained on any ballot in the election including special meaning marks such as party line vote. Columnar tabulation generally results in an election result. When the CVRs are appropriately sorted by a subset such as by precinct, then subtallies by precinct can be easily aggregated.

The structure of the cast vote record ought to permit an optional extension that represents a figure of merit for each mark such as a mark density within target or another vendor defined metric. This could appear in a separate CSV table with similar layout.

The header row of the CVR should contain identifying codes that unambiguously identify the contest and the choice represented by each column. These could be numerical codes that are identified in a separate election specific XML document by corresponding UTF-8 strings or alternatively, if the storage is efficiently used, each UTF-8 string could be contained in the corresponding fields of the first row of the CSV table.

Note that it is crucial that the identifiers for contests and choices be identical across entire districts and not just specific to election jurisdiction so that statewide processing of CVRs can be accomplished. Thus whether or not numerical identifiers are used, a statewide authority must allocate these identifiers and enforce that they remain identical when jurisdictions are sharing contests for an election district. States encounter the same problem with election night reporting and so the solutions may be similar or the same.

Note also that some CSV fields may have to be left empty to preserve the anonymity of the ballot when exposure of the equipment used or the origin or voting method of the ballot is so rare that the style itself within that subset identifies the voter of the ballot.

Note also that it is highly desirable that the set of CVRs be encapsulated for an entire election within one consistent file (or two or three as proposed here). Some initial efforts at producing cast vote records have produced one file per ballot or one file per batch. Too many files is a formula for lost records.

There is also a justification to have a digital signature calculated for the CVR files so that equality with the approved set can be immediately ascertained. Officials such as auditors or members of the public who may find inconsistencies using the cast vote records when reporting should be asked to provide the source materials used so that accuracy of the copy can be determined using the digital signature.

What follows in the form of an appendix is the new Colorado Election Rule for the Cast Vote Record. I include also my comments in similar form to that submitted to the CO SOS regarding the details of this rule as proposed. I do not recommend all of the specifics of the promulgated rule, as shown by my inline comments, largely shown in blue.

21.4.14 Ballot-level Cast Vote Records and Exports. All voting systems certified by the Secretary of State for use in Colorado on or after January 1, 2016 must meet the following requirements for ballot-level cast vote records and exports on or before December 31, 2016:

- (a) The voting system must capture a ballot-level cast vote record (CVR) consisting of a single record for each ballot tabulated, showing the manner in which the voting system interpreted and tabulated the voter's markings on the ballot, as adjudicated and resolved by election judges, if applicable.
- (b) The voting system must be able to aggregate in a single file and export all CVRs in comma-separated value (CSV) text format.
- (c) The CVR export must contain the following fields, with values or data populated by the voting system:
 - (1) CVR Number. A sequential number from one to the number of CVRs in the export file. This can be used as an alternate method to identify each CVR.
 - (2) Batch ID. Identifies the batch in which the paper ballot corresponding to the CVR is located.
 - (3) Ballot Position. Identifies the position of the paper ballot corresponding to the CVR within the batch.
 - (4) Imprinted ID. If the scanner model supports imprinting a unique character string on the ballot during the scanning process, the voting system must populate this field with the unique character string.
 - (5) Ballot Style. Indicates the ballot style of the paper ballot corresponding to the CVR.
 - (6) Device ID. Identifies the scanning device by model, serial number, and/or scanning station identifier.
 - (7) Contest and Choice Names. Each contest and choice on any ballot in the election must have its own field so that voters' choices in all contests can be easily and independently tabulated after the CVR export is imported into a spreadsheet application.
- (d) The header or field names in the CVR export must unambiguously correspond to names of the contests and choices on the paper ballots.
- (e) The contests and choices must be listed in the same order as they appear on the ballots.
- (f) A vote for a choice must be indicated by a "1". No vote for a choice or an overvoted condition must be indicated by a "0". Choices that are not applicable to the CVR must be left blank.

Following FYI is an inline commentary to the CO SOS rulemaking process that produced the above rule.

Harvie Branscomb Jan. 13 2016 comments on CO SOS Elections Rulemaking concerning Cast Vote Record

New Rules 21.4.14

21.4.14 SINGLE-BALLOT CAST VOTE RECORDS AND EXPORT. ALL VOTING SYSTEMS SUBMITTED FOR CERTIFICATION AFTER DECEMBER 15, 2015 MUST COMPLY WITH THE REQUIREMENTS OF THIS RULE 21.4.14.

(A) THE VOTING SYSTEM MUST CAPTURE A CAST VOTE RECORD (CVR) CONSISTING OF A SINGLE **BALLOT** SET OF ENTRIES FOR EACH BALLOT TABULATED, SHOWING THE **MANNER IN WHICH RESULT OF** THE VOTING SYSTEM INTERPRETATION ~~ED~~ ~~AND TABULATED~~ OF THE VOTER'S MARKINGS ON THE BALLOT AS ADJUDICATED IF REFERRED TO ELECTION JUDGES..

(B) THE VOTING SYSTEM MUST BE ABLE TO AGGREGATE IN A SINGLE FILE AND EXPORT ALL CVRS **FOR ONE ELECTION** IN COMMA-SEPARATED VALUE (CSV) TEXT FORMAT.

(C) **EXCEPT THOSE MARKED AS OPTIONAL,** THE CVR EXPORT MUST CONTAIN THE FOLLOWING FIELDS, WITH VALUES OR DATA POPULATED BY THE VOTING SYSTEM:

(1) CVR NUMBER. A SEQUENTIAL NUMBER FROM ONE TO THE NUMBER OF CVRS IN THE EXPORT FILE. THIS CAN BE USED AS AN ALTERNATE METHOD TO ~~IDENTIFY~~ **ASSOCIATE EACH CVR WITH THE RELATED PAPER BALLOT AND ANY OTHER DIGITAL IMAGE RELATED TO IT.**

(2) BATCH ID. **(COUNTY OPTIONAL)** IDENTIFIES THE BATCH IN WHICH THE PAPER BALLOT CORRESPONDING TO THE CVR IS LOCATED. **THIS FIELD IS NOT APPROPRIATE FOR A COUNTY IN WHICH BATCHES BY NECESSITY OR BY CHOICE SIGNIFY AN IDENTIFIABLE GROUP OF VOTERS SUCH AS AT A SINGLE VSPC. IN SUCH CASE, THIS FIELD MUST BE SUPPRESSED FROM THE CVR. COUNTIES PERFORMING 100% CENTRAL COUNT MAY INCLUDE THIS FIELD WITHIN THE CVR.**

[In such counties the ballot manifest that contains the batch and location information can be kept separately so that the CVR can be committed to as a public record. In most but perhaps not all cases, the batches can be crafted so that they do not correspond to a known list of voters. When that is not possible, the suppression of the batch number from the CVR is needed.]

(3) BALLOT POSITION. IDENTIFIES THE POSITION OF THE PAPER BALLOT CORRESPONDING TO THE CVR WITHIN THE BATCH. ~~TARGET CARDS~~ ~~SCANNED TO IDENTIFY THE BATCH MUST NOT BE INCLUDED IN THIS~~ ~~COUNT.~~

(4) IMPRINTED ID. IF THE SCANNER MODEL SUPPORTS IMPRINTING A UNIQUE CHARACTER STRING ON THE BALLOT DURING THE SCANNING PROCESS, THE VOTING SYSTEM MUST POPULATE THIS FIELD WITH ~~THE A~~ **UNIQUE SEQUENTIAL CHARACTER STRING THAT IS NOT ASSOCIATED WITH A VOTER.**

(5) BALLOT STYLE. INDICATES THE BALLOT STYLE OF THE PAPER BALLOT CORRESPONDING TO THE CVR.

(6) **DEVICE ID. (COUNTY OPTIONAL)** IDENTIFIES THE SCANNING DEVICE BY MODEL, SERIAL NUMBER, AND/OR SCANNING STATION IDENTIFIER. **THIS FIELD IS NOT APPROPRIATE FOR A COUNTY IN WHICH DEVICES BY NECESSITY OR BY CHOICE SIGNIFY AN IDENTIFIABLE GROUP OF VOTERS SUCH AS AT A SINGLE VSPC. IN SUCH CASE, THIS FIELD MUST BE SUPPRESSED FROM THE CVR. COUNTIES PERFORMING 100% CENTRAL COUNT MAY INCLUDE THIS FIELD WITH THE CVR.**

(7) **CONTEST NAME AND ID.** EACH CONTEST ON ANY BALLOT IN THE ELECTION MUST HAVE ITS OWN FIELD. THIS FIELD CONTAINS EITHER A "V" FOR FULLY VOTED, A "U" FOR UNDERVOTED IN A SINGLE CHOICE CONTEST, A "P" FOR UNDERVOTED IN A MULTIPLE OUTCOME CONTEST, AN "O" FOR OVERVOTED, A "W" FOR ONE OR MORE WRITEINS OR A BLANK FOR A CONTEST NOT ON THE BALLOT. THESE ENTRIES MAY BE USED TO SORT THE CVRS FOR VERIFICATION PURPOSES.

(8) **CONTEST AND CHOICE NAMES AND IDS.** EACH **CONTEST AND CHOICE INCLUDING WRITE-INS ON ANY BALLOT CAST** IN THE ELECTION MUST HAVE ITS OWN FIELD SO THAT VOTERS' CHOICES IN ALL CONTESTS CAN BE EASILY AND INDEPENDENTLY TABULATED AFTER THE CVR EXPORT IS IMPORTED INTO A SPREADSHEET APPLICATION. **THESE FIELDS MAY CONTAIN EITHER A "1" OR A "0" OR BE LEFT BLANK.**

(9) **(OPTIONAL) FOR EACH CONTEST CHOICE AN ADDITIONAL ADJACENT FIELD THAT CONTAINS THE MARK DENSITY OR OTHER MEASUREMENT OF THE MERIT OF THE VOTER MARK DETECTED BY THE VOTING SYSTEM.**

[Having this in the CVR will greatly improve our ability to audit the election data starting from the most likely misinterpreted voter marks.]

(D) THE HEADER OR FIELD NAMES **FOR CONTEST AND CHOICE FIELDS** IN THE CVR EXPORT MUST ~~UNAMBIGUOUSLY~~ CORRESPOND **UNAMBIGUOUSLY FOR ANY OBSERVER** TO NAMES OF THE CONTESTS AND CHOICES ON THE PAPER BALLOTS. THE USE OF AN **ALPHANUMERICAL CHOICE ID AND CONTEST ID TO IDENTIFY SIGNIFY EACH CONTEST AND CHOICE MUST BE AVOIDED BECAUSE THEY REQUIRE CROSS-REFERENCING TO OTHER SOURCES TO DETERMINE THE CHOICE AND CONTEST NAMES. THIS FIELD IS ALSO NECESSARY IN ORDER TO SIMPLIFY THE USE OF THE CVRS FOR AUDIT OF MULTICOUNTY CONTESTS. THE IDS SHOULD MATCH THOSE USED TO COMPLY WITH RULE 21.4.15.**

(E) THE CONTESTS AND CHOICES MUST BE LISTED IN THE SAME ORDER AS THEY APPEAR ON THE BALLOTS.

(F) A **MARK SIGNIFYING A TABULATED** VOTE FOR A CHOICE MUST BE INDICATED BY A "1". **MARKS THAT PRODUCE NO VOTE FOR A CHOICE OR AN OVERVOTED CONDITION** MUST BE INDICATED BY A "0". CHOICES THAT ARE NOT APPLICABLE TO THE ~~CVR~~ **VOTER OR CONTESTS FOR WHICH THERE IS NO MARK FOR ANY CHOICE (UNDERVOTE)** MUST BE LEFT BLANK. **FOR EVERY INTENTIONAL MARK FOR A CHOICE ON EACH PAPER BALLOT THE CVR SHOULD CONTAIN EITHER A "1" OR A "0".**