Preserving Anonymity of Cast Vote Records

Mark Lindeman, John McCarthy, Neal McBurnett, Harvie Branscomb, Ron Rivest, & Philip Stark, 8/3/2017

Executive Summary

The use of cast vote records can greatly increase the efficiency of audits, but care must be taken during the creation of cast vote records and during the audit to ensure that they cannot be linked to particular voters.

Colorado has passed a law requiring a "risk-limiting audit" (RLA) to be conducted starting in 2017. An RLA checks the accuracy of vote tabulation: specifically, it seeks strong evidence that election outcomes match those that a full hand recount would obtain. The most efficient RLAs make use of cast vote records (CVRs), which report the machine (voting software) interpretation of the voter's selections on each ballot. The audit proceeds by examining a random sample of paper ballots, manually interpreting the votes on those ballots, and comparing the corresponding manual and machine interpretations. (Accordingly, this method is called a ballot-level comparison audit.) Fifty-four of Colorado's 64 counties have acquired equipment capable of creating these CVRs as of 2017.

Colorado's elections combine many multifariously intersecting districts (legislative, county, municipal, school, special), producing a profusion of ballot styles. About 95% of Colorado's ballots are distributed and returned in the form of mail ballot packets. Upon return, unsorted envelopes containing ballots are divided into batches for eligibility determination. An identifiable return envelope associated with a specific residence address and ballot style is checked for evidence of eligibility.

Colorado law requires secrecy in voting. This requirement entails that ballots being prepared for tabulation must be separated from the identity of the voter. Because CVRs summarize the content of ballots, they can pose similar threats to secrecy as ballots themselves. And because CVRs are crucial to efficient audits, they must be available to the public so that these audits are observable and verifiable. Therefore, it is crucial that each CVR also be systematically separated from the associated voter identity revealed on the envelope.

This document **addresses potential risks** for correlation of the CVR with the identity of the voter, and offers possible means of preserving the **anonymity of the CVR** so that it may be used as required for the audit. Because of the close relationship between CVRs and ballots, concerns related to CVRs often focus attention on procedures that inherently jeopardize ballot anonymity as well. The best solutions address both kinds of problems. Some of these solutions (e.g., changing how district boundaries are drawn) are beyond local election officials' control. Adopting best practices for batch redefinition, as in draft election rule 7.10.5, is an important way to address threats to CVR and ballot anonymity. CVR partial-redaction methods and RLA-specific remedies can serve as stopgaps, although these approaches have significant shortcomings and disadvantages.

.

Principles:

1) Audit process and materials must be available for in-person public inspection and verification
2) Public access to physical ballots is necessarily limited by logistics, so alternative means for remote access to credible copies of the evidence used in audits are needed
3) The evidence (data) to be audited must be "committed to" in public before generating the random seed for the audit. Commitment entails, at a minimum, publishing a hash value calculated from the data, and subsequently publishing the data so that anyone can use the hash value to confirm that the published evidence is identical to the evidence used in the audit.
4) Distinct levels of access are applicable to specific categories of individuals
    a. Election officials (clerk, staff, SOS staff, canvass board, contractors)
    b. Watchers (credentialed by interested parties)
    c. Candidates
    d. Public
5) Modes of access to evidence may vary depend on the type of evidence:
    - Original voter marked paper ballot
    - BMD printed ballot voted in person
    - BMD printed ballot voted remotely
    - Duplicated ballot representing original voter markings
    - Image of ballot as scanned by voting system
    - Photograph of voter marked paper ballot taken by auditors
    - Instances of above categories where voter has self-identified
    - Cast Vote Record as emitted by the voting system for unofficial results
    - CVR as uploaded and published for audit
    - CVR as updated during audit

6) Counties are required to cover or redact certain "voted ballots" before "making available for public inspection" under Colorado Open Records Act (C.R.S.24-72-205.5):
    - If voter has self-identified through marking
    - If there are fewer than 10 instances of a ballot format or portion thereof in the election

Discussion:

Anonymity issues affect access to paper ballots, scanned images, and photos of paper ballots, but these topics are not addressed here. Risks caused by voter self-identification of the paper ballot and copies thereof are not under discussion. Self-identification of ballots via pattern voting is also not discussed, as it cannot be used to provably identify a ballot or a CVR and requires the cooperation of the voter. Self-identification with paper ballots and images thereof is currently allowable under Colorado law.

Here we are concerned with the potential for a published or viewed CVR to be identified with a particular elector, and the steps that must be taken to largely remediate this risk.

The susceptibility of the CVR to identification with the voter depends on the policies and procedures of each county that uses a voting system that is capable of creating a CVR. It also depends on the fields chosen to be included in the CVR.

In Colorado, the definition of the CVR is still somewhat dynamic. It is subject to recent rulemaking from 2016 and 2017. The Appendix contains the CO definition of the CVR under Election Rule 21. There will likely be 8 fields required by rule to be included.

If counties follow proposed Colorado Election Rule 7.10.5 properly, there are a small number of situations which still need to be addressed, which we consider under (A). If they do not follow the processes described in that rule, the situation is much more problematic, which we consider as case (B).

(A) The **envelope batches are disassociated from the tabulation batches** physically and logically prior to tabulation per proposed Colorado Election Rule 7.10.5:

In this case the only CVRs that will be identifiable are those for which the style plus any other discriminating factor (e.g. “CountingGroup,” discussed below) is unique within any collection that can be tied to a set of voters – that is, any collection that matches a group of ballots that is maintained intact (or almost intact) from the eligibility check to tabulation. In small counties, it is likely that the only such collection is the set of all ballots in the county. In larger counties, it might represent a tub of 1000 or 2000 ballots. Harvie Branscomb has elsewhere (see link below) proposed requiring that any such collection contain at least five times as many ballot pages as the number of styles contained within. This “collection” represents an intentionally large scale partitioning of the ballots at which direct accountability of counts of pages is maintained.

http://www.sos.state.co.us/pubs/rule_making/written_comments/2017/20170718Branscomb.pdf

Under proposed rule 7.10.5 there remain three cases of anonymity risk at the CVR level built into the current Colorado election system that must be watched for. Once these are remedied, the CVRs can be published without risk to voter privacy.

1) Rare ballot styles

In some counties there are unfortunate cases where a coordinated ballot style is printed in very small quantity, bordering on unique. Instances exist in which perhaps only one or two ballot pages of these rare styles are cast. Such cases arise from, e.g., a school district border that briefly crosses a county line, or a congressional district that does so (e.g. CD1 voters in Jefferson County in a previous decade). Because publicly available records reveal which voters are eligible to cast these rare ballot styles, and which of them voted, both the CVRs and the ballots themselves ineluctably threaten secrecy. These CVRs and ballots should not be published or publicly audited. In fact, district lines that create rare ballot styles should not be allowed, since they inherently compromise the constitutional guarantee of secrecy in voting.

Fortunately, risk-limiting audit theory allows for the handling of such rare and inaccessible ballots. They can be treated as if they are voted in a manner that least confirms the presumed winning choices (i.e., the worst case). The existence of these ballots tends to add to the number of discrepancies and increases the workload of the audit, perhaps only enough to motivate an eventual solution to the source of the rare style.

These cases are rare and trivially soluble at the time the setting of district boundaries is done. Another opportunity for remedy occurs before the county agrees to coordinate to add the unfortunately designed district onto a county ballot via an Intergovernmental Agreement. At both times this problem can be solved by following best practices that remain insufficiently understood.

One such solution is to design the coordinated election to isolate the contest on a separate ballot, or equally effectively, to leave it in a separate election. This link leads to a proposal for change in the Form of Ballot statute to encourage a smart separation of contests on multiple ballot pages in a coordinated election.

http://www.sos.state.co.us/pubs/rule_making/written_comments/2017/20170718BranscombMcCarthy.pdf

A possible stopgap solution is to modify references (in the Dominion CVR field named "BallotType") to ballot styles rendered rare by specific contests and redact those specific contest columns of the CVR to remove only the offending district election (e.g. school board) for all CVRs. It may be wise to redact similarly the CVRs in all other affected counties (to avoid the mistake of presuming that all ballots containing the contest are represented while a county contribution to the results is missing). This approach does not solve the anonymity problem with any paper ballot to be audited, and makes an assumption that the risk to anonymity of fleeting access to the paper is minor and that the problematic contest will not be captured into the record of the audit. To implement such a redaction, not only do the choice columns for the problematic contest need to be removed from all affected county CVRs, but the style numbers pertaining to rare rows need to be edited so that these newly styled CVR rows merge with others not containing the offending contest. Note that this modification to the CVR is not trivial and may result in unanticipated errors and cannot be verified by the public. These represent two major disadvantages that need not be experienced if the other two alternatives are pursued (not coordinating and properly setting district borders).

2) *Unusual voting method*

The Dominion cast vote record contains a column entitled "CountingGroup," which undesirably reveals evidence pertaining to whether the ballot was cast in-person and whether a Ballot Marking Device was used. This evidence is also revealed to the public in the "who voted" list that parties use. Considering that only a few percent of voters choose to vote in-person in Colorado, the number of those voters using each style of ballot is very low. Thus, this information compounds the problem of rare ballot styles. Many of the in-person ballots and CVRs, if they include the CountingGroup field, are rare enough to require treatment as anonymity-risk ballots. Even after the CountingGroup field is removed from the CVR, such BMD/QR ballots may be, by non-ideal procedure, grouped in specific batches or planned to be counted on a specific scanner, thus revealing their category (in effect, CountingGroup) in the scanner ID. Anonymity is best ensured by mixing these ballots in with those from other voting methods as they are scanned. This is one more reason to re-create batches prior to tabulation.

3) Military and overseas (UOCAVA) ballots

Military and overseas (UOCAVA) ballots can be treated much like in-person ballot marking device ballots. In this case, however, the voter signs an affidavit recognizing the built-in lack of privacy of the transmission method for the ballot. The so-called UOCAVA ballots are even more rare than in-person ballots, but as far as one can predict, these ballots will not be separately identified in the CVR, unless by batch or scanner ID if the county decides to use a specific scanner or a special set of batches only for UOCAVA. UOCAVA batches might be indirectly identifiable because they contain Federal Write-in Ballots that can be recognized via CVR.

In conclusion, when ballots have rare styles – say, if nine or fewer ballot pages of that style are printed and sent out (a higher number would be preferable) – counties must be ready to avoid auditing these ballots if they are randomly sampled, and treat them as exceptions. To defend anonymity of in-person BMD ballots, from a CVR point of view, the CountingGroup field must be removed from the Dominion CVR format. For UOCAVA voters, the original ballot tends to be identifiable even if the duplicate ballot is not. Among these UOCAVA ballots, those that have identifiable voter verified originals that were subsequently duplicated for tabulation, if randomly sampled, should also be considered inaccessible and treated as exceptions.

(B) In this alternative situation, the procedures in proposed Colorado Election Rule 7.10.5 are not followed. Thus the **envelope batches and tabulation batches are closely related** in content (only a few ballots are either missing or removed for ineligibility reasons, and these are carefully accounted for in batch tracking records). Some measures may have been taken to improve anonymity of ballots in batches, such as to shuffle the order of ballots within batches, and the envelope batch numbers may have been textually disassociated from and no longer relate to the tabulation batch numbers. The following concerns remain after these measures have been taken:

Many of the fields contained within the CVR still give ample evidence about the match between the envelope and tabulation batches. Batch sizes are likely small compared to the number of styles printed (including in some elections, precinct styles). Once batches are matched, all of the unique styles contained will be identifiable. In many more instances of rare styles, the voter intent will be exposed for all voters in the same style if they vote alike in a contest.

Here are examples of ways to use various CVR fields to match disassociated batches from eligibility check (envelopes are identifiable) to tabulation (ballots presumably anonymous):

Batch ID: Use of a common Batch ID between eligibility and tabulation offers an easy means to match the two. Once renumbered, the count of batches and the order of numbered Batch IDs may provide evidence for matching the batches. Redaction of the BatchID is not a solution because the BatchID (and at least indirectly, container ID) is needed to locate the paper ballot to match to the CVR for the audit.

Ballot Position: the maximum of this sequence number will indicate the size of each batch. The count of ballots in a batch is very good evidence for matchup of batches used for envelopes and used for tabulated ballots. Making all batches the same size could resolve this issue, but because of unexpected changes in batch size after envelope opening due to missing and double ballots returned, equal size batches are difficult to obtain without re-batching. Again, either Ballot Position or Imprinted ID will be needed to locate the ballot within a batch. One of these fields is needed within the CVR to accomplish the audit.

Imprinted ID: Assuming that this is a semi-sequential number to aid in ballot location, the range of Imprinted IDs will also reflect the batch size and possibly batch tabulation order. If these are sequential across batches, they will also match loosely to envelope batch creation order, thus revealing clues to batch identity.

BallotType (as named in Dominion's CVR format): This field contains the Ballot Style as defined in rule. The frequency of various ballot types (styles) in a batch almost certainly uniquely identifies it to a specific envelope batch. That is true unless envelopes were sorted by style (or sets of styles) prior to batching, so that the ballots within a batch are provided anonymity by uniformity in style. Ballot style is inherent to the CVR for whatever contest choice fields are included. These could be redacted at the loss of auditability for those contests, but the Ballot Style entry for each CVR row that contains redaction would have to be edited to prevent it from revealing the redacted contest within the style number.

These matching methods are possible because of the dangerous policy of maintaining batch identity from envelope to tabulation. How do we redact the CVR in order to prevent use of the above methods of matching? The remaining fields within the CVR must be adequate to permit public commitment and sharing of at least the CVR portion of the election record. Far better would be a sharing of access to the paper and a sharing of the ballot scanned images, but that is beyond the scope of this discussion.

One method proposed by Neal McBurnett would suffice to defend voter privacy and allow public access to the audit assuming that the audited contest(s) are all countywide. That method would simply split the CVR such that remaining contests reported are all countywide or greater. This means removing all columns that relate to partial county contests. At the same time, the column for Ballot Style (BallotType) must also be removed along with "CountingGroup." This would put into effect a simulation of a two-card ballot election where the second card contains only the non-countywide contests and it is not audited, or at least, in this case, ignored for purposes of audit. The virtual second card is separated from the first card, eliminating any rare style association with the first card. This approach removes the style consciousness of the CVR, making the CVR amenable to delivery to the public including under the terms of the Colorado

Open Records Act (CRS 24-72-205.5 (4)(b)(III)). And it paves the way for an adoption of a two-card ballot anonymity defense that isolates all the style complications to a separate physical card, tabulated separately and producing a separate CVR.

If there is no countywide contest, and the SOS has selected a partial county contest for “audit,” then the redaction must be more sophisticated or more drastic in order to permit an audit of some contests to be held in public. For example, only the entries for a particular set of contests could be committed to and published in a CVR file without the columns for “CountingGroup” or “BallotStyle.”

Obviously these redaction approaches need extension when a combination of countywide and partial county contests are being audited – as we expect for the 2018 election. In that case, it is crucial to consider the number of unique styles in any collection that matches an associated collection of envelopes, to see if criteria for voter privacy are satisfied. At present those criteria are only available in Colorado’s CORA statute, 24-72-205.5. It would be better to have election-auditing-specific criteria for maintaining voter privacy, and particularly criteria that solve the problems as close to the source as possible. A very useful step for solving these issues is sorting of ballots by style (or groups of styles), either before envelope batches are formed, or after opening envelopes but before forming tabulation batches. Sorting ballots by machine or by hand, according to style or groups of styles, while the ballots are still in envelopes is beneficial if envelope batches will continue to match tabulation batches.

APPENDIX

Excerpt of Election Rule 21

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.

(8) {Proposed} ***NUMBER OF VALID CHOICES.***

THE NUMBER OF VALID CHOICES (E.G., "VOTE FOR 3") FOR EACH CONTEST.

(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.

Constitutional provision for anonymity (secrecy in voting).

> **Colo. Const. Art. VII, Section 8** ELECTIONS BY BALLOT OR VOTING MACHINE
>
> *All elections by the people shall be by ballot, and in case paper ballots are required to be used, no ballots shall be marked in any way whereby the ballot can be identified as the ballot of the person casting it. The election officers shall be sworn or affirmed not to inquire or disclose how any elector shall have voted. In all cases of contested election in which paper ballots are required to be used, the ballots cast may be counted and compared with the list of voters, and examined under such safeguards and regulations as may be provided by law. Nothing in this section, however, shall be construed to prevent the use of any machine or mechanical contrivance for the purpose of receiving and registering the votes cast at any election, provided that secrecy in voting is preserved.*

Relevant statute. Note that Colorado Open Record Act is typically relied upon to define access to election records for all persons other than election officials. These others include the public who might attend the audit or attempt to verify its process and conclusions. Note that "internal batch reports … for the purpose of auditing" are specifically excluded from access during the "stay period" prior to certification deadline. This sentence was included over strong objections at the time (2012). Some language in CRS 24-72-205.5 says "before the ballot may be made available for public inspection" and that statement does not appear to be dependent on a prior CORA request. Indeed it is unreasonable to expect the access to "ballots" during the audit to be subjected to limitations that apply for all members of the public attempting to gain access to public records. But it may be seen as reasonable for these restrictions to apply to records that are about to be published for the purpose of verification. Note then that the word "ballot" refers to the ballot scan as well as the Cast Vote Record.

> **C.R.S. 24-72-205.5 (2)(a)** *"Ballot" means a ballot voted by any acceptable, applicable, or legal method that is in the custody of an election official. "Ballot" includes any digital image or electronic representation of votes cast.*
>
> **C.R.S. 24-72-205.5 (3) (a)** *Except as otherwise provided in paragraph (b) of this subsection (3), the designated election official shall not fulfill a request under this part 2 for the public inspection of ballots during the period commencing with the forty-fifth day preceding election day and concluding with the date either by which the designated election official is required to certify an official abstract of votes cast for the applicable candidate contest or ballot issue or ballot question pursuant to section 1-10-102 or 31-10-1205 (1), C.R.S., as applicable, or by which any recount conducted in accordance with article 10.5 of title 1, C.R.S., or section 31-10-1207, C.R.S., is completed, as applicable, whichever date is later. The denial of public inspection of ballots authorized pursuant to this paragraph (a) shall also apply to any internal batch reports generated by a designated election official for the specific purpose of auditing ballots received in the course of conducting an election.*

C.R.S. 24-72-205.5 **(4) (b) (I)** *The original ballots shall at all times remain in the custody of the designated election official or his or her designee. In the discretion of the designated election official or his or her designee, and subject to the provisions of paragraph (a) of this subsection (4) and this part 2, the designated election official or his or her designee shall determine the manner in which such ballots may be viewed by the public.*

(4) (b) (II) The designated election official or his or her designee shall cover or redact, based upon the most practical means available, any markings or message on a ballot that may identify the particular elector who cast the ballot before the ballot may be made available for public inspection;

(4) (b) (III) To protect the privacy of particular electors, any ballots cast by electors within groups of discrete individuals who are more susceptible of being personally identified, such as military and overseas electors, shall be made available for public inspection only to the extent such ballots may be duplicated without identifying elector information. Insofar as such ballots are not able to be duplicated without identifying elector information, they are not available for public inspection. Notwithstanding any other provision of this section, no ballot, or any portion thereof, may be made available for inspection where the ballot, or any requested portion thereof, is identical in printed form, considering a combination of the election contests at issue and precinct coding, to only nine or fewer ballots, or comparable portions thereof, among all ballots used in the same election. However, any such ballot, or any requested portion thereof, that is identical in printed form to ten or more ballots, or comparable portions thereof, used in the same election may be inspected.

(4) (b) (IV) To protect the privacy of particular electors, ballots made available for inspection may be presented in random order selected by the designated election official or his or her designee;

…

(5) Notwithstanding any other provision of this section, nothing in this section affects either the rights of a watcher set forth in the provisions of titles 1 and 31, C.R.S., or the operation of a canvass board in accordance with the provisions of articles 1 to 13 of title 1, C.R.S.

Sample of Dominion Cast Vote Record as provided by Dwight Shellman

(Many other columns on the right hand side of the figure, representing choices in various contest choices, have been redacted for clarity and column names have been split onto two lines for readability. Also, the CVR file has been sorted by "BallotType.")

								Presidential Electors (Vote For=1)	Presidential Electors (Vote For=1)
								Hillary Clinton / Tim Kaine	Donald J. Trump / Michael R. Pence
Cvr Number	**Tabulator Num**	**Batch Id**	**Record Id**	**Imprinted Id**	**Counting Group**	**Precinct Portion**	**Ballot Type**	**DEM**	**REP**
128	9	3200	49	9-3200-49	Mail	4295603363-01 (363-01)	1	1	0
129	9	3200	50	9-3200-50	Mail	4295603363-01 (363-01)	1		
156	10	2	16	10-2-16	In Person	4295603363-01 (363-01)	1	1	0
93	9	3200	3	9-3200-3	Mail	4295603365-02 (365-02)	2	0	0
130	9	3200	4	9-3200-4	Mail	4295603365-02 (365-02)	2		