

From: Harvie Branscomb [REDACTED]
Sent: Wednesday, July 29, 2015 4:19 PM
To: Dwight Shellman
Cc: [REDACTED] McReynolds, Amber Faye - Elections Division; Matthew Crane
Subject: are we ready for risk limiting audits?

Dwight Shellman

Here is an excerpt of something I am sharing with election verification experts:

Fortunately as Phil Stark pointed out ballot level risk limiting audits do not require aggregating multiple ballots and only actually require capture and comparison of patterns off of single ballots at a time. However this is also not trivial and even newest systems are not necessarily providing data that has good human factors for this purpose.

It appears that machines are better than humans at counting up the number of pieces of paper in front of them. It seems unavoidable that machines fail to capture idiosyncratic voter intent when humans succeed. Therefore an ideal combination is found in the sort and stack method that can be combined with machine counting of pages. For example during a single contest recount it would be ideal to have human judges separate the ballots by voter choice and batch tabulate them such that all ballots in a batch (or on a particular scanner or memory card) are hand sorted to be a single voter choice. The scanner can then provide a machine count of sorted pages and a tally of vote counts to confirm that it recognized all as voted for the specific choice. This will easily expose discrepancies that can be resolved by human observation and improve recount accuracy.

I am currently analyzing the first 8,000 of 130,000 ballot images and cast vote records from the Dominion system used in Denver CO (the much advertised and visited May municipal pilot election). The principal reason for my analysis is to learn what we need to know to guide the evaluation of such a pilot system from the point of view of auditing (and transparency)... auditing the audit as it were.

The format for the Cast Vote Records that were provided 1) as a separate text file aggregated by batch as well as 2) the "AuditMark" voter selection digital printout that is embedded with the 200dpi 1bit per pixel ballot image are not suitable for convenient machine tabulation and they aren't particularly desirable for hand comparison either. Human usability concerns (as well as machine formatting compatibility concerns) abound in the design of forms and scanned images and reports that are needed for auditing. No doubt this issue is not limited to Dominion.

Our CO pilot review program seems insufficiently concerned about the ways that the risk limiting audit will take place. Perhaps more attention will be paid within the next few weeks. We

have until November when five such pilots will take place as part of the 2015 election. At present it is uncertain what the instructions for a risk limiting audit will be and it is even more uncertain how we will know if we have done a good pilot audit.

Dwight I am deeply concerned that our pilot evaluations will fail to measure and document the crucial back office aspects of the systems even as they are in use in real elections. The evaluations of these systems in real elections crucially must look beyond the motions of successfully performing an unremarkable election. The RLA is a big part of my concern and that is why I asked you to forward the communication to the vendors about preparations for the pilots. Should I ask for that again?

I am hoping that I will get enough information from Denver's completed pilot in time to be able to make a sample partial evaluation of that system as an example for what the PERC or SOS could and should do as a minimum to evaluate each system. I am hoping I can do that before the PERC decisions on how and what to evaluate are made. No doubt what I am doing by myself is not enough either. I am already learning that the RLA may be painful with some systems. And the counties performing the pilots may have the tendency to act to defend the system rather than subject it to a critical evaluation in public. That should be a major concern of PERC and SOS and all Coloradans.

I don't think my bi-weekly writings to the PERC have been productive but I do think that you can benefit from the information I am collecting. By random chance after examining only a few ballot images in Denver I already found an incorrectly captured voter mark in the Dominion May pilot- but the CO evaluation must be in a position to find all of them for each piloted system. And system evaluators must be in a position to understand and evaluate the remedies offered by the vendors. A technique such as I have briefly described above for sorting before scanning if used for a top of ticket contest would get part of the way there and without spending much extra time.

Likewise the documentary products of each system (logs, settings, etc.) must be studied in detail and each document's ability to convey exception cases and discrepancies evaluated. Unfortunately I have not yet received any logs or reports from the Denver system, and the ballot images are coming slowly due to an understandable learning curve on privacy redaction. The timing of the PERC and any other SOS evaluation must allow for weeks of evaluation of reports and cast vote records and audits after the documents are produced. It will not be enough to say no one complained.

And on that matter, it is also important that well trained watchers/observers be present at the 10 pilots at important moments throughout - of course including the ballot design, LAT and ballot prep and early tabulation and post election activities. And plans must be made for a super-LAT at each pilot that plans for potentially problematic ballots to be voted and results analyzed. Remember what the clerks did to ballots for the UVS demo last year- crumpled them, etc. Whatever the out of state certification is doing and whatever SOS is doing in house is probably insufficient to provide confidence about exceptional test cases and the ability of systems to detect, log and remedy these cases.

Finally you keep mentioning the hardware test and acceptance tests. These are important and have been ignored in the past. I think work should be done on what the hardware tests (every election) and acceptance tests (by county upon purchase) should be- and these tests should be performed with careful evaluation during the pilot or within its scope. Seeing the lights come on and the system say "ready" doesn't provide enough information.

Please send this to the PERC if you feel it is constructive to do so.

PS I just received the following link from Ron Rivest- another very well respected statistician working on elections - with a simple one page description of a RLA. Unfortunately it is still only the mathematicians description and not one for election officials. Someone should turn this (or similar mathematical instructions) into a clear and specific multi-county process (all 5 RLA pilots coordinated plus any other counties separately

doing RLA?) for the state-wide issue on the ballot and within county audits for the remaining contests.

<http://people.csail.mit.edu/rivest/pubs.html#Riv15y> In this list of publications, the RLA description is number 311 or the second from the top at the moment.

Harvie Branscomb