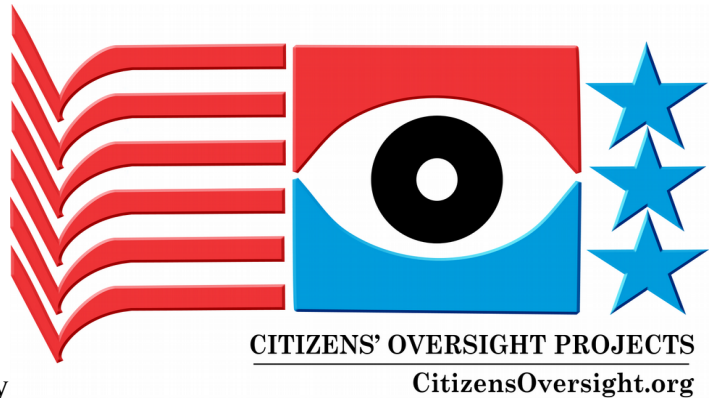


## Citizens' Oversight Projects (COPs)

771 Jamacha Rd #148  
El Cajon, CA 92019  
CitizensOversight.org  
619-820-5321

May 28, 2018

Assembly Elections & Redistricting Committee  
c/o Lori Barber, Consultant & Committee Secretary  
1020 N Street, Room 365  
Sacramento, California 95814  
916.319.2094  
lori.barber@asm.ca.gov



COMMENT ON AB-2125 “Risk Limiting Audits”  
Version: “Amended in Assembly May 21, 2018”

This is a much better version of the bill than the one first proposed. I thank you for the improvements, but we are not comfortable with this version for the following reasons:

1. Although this audit process does provide a great final check on the election, it does not provide sufficient direction and recommended procedures for election officials to employ to check on the quality of the canvass as it is being processed. The guideline for all quality assurance testing is “test early and test often”. The Risk Limiting Audit is only a final check and does not provide an adequate level of confidence for election officials during the process and it certainly does not provide feedback for early intervention should difficulties arise.
2. This version puts great dependency on the Secretary of State to “define in regulations the vote totals to be used in the comparison audit.” Apparently, that means that the Secretary of State can decide to short-cut the audit and use only a fraction of the ballots returned. We assert that this final statistical check should include all ballots in scope of the sampling process. Sampling procedures used to provide a statistical check on the results must not ignore large groups of ballots, such as the “Later Vote-by-Mail” (those not fully processed until after election day) and accepted provisional ballots.

The methodology used in statistical sampling is called “stratification.”

In statistical surveys, when subpopulations within an overall population vary, it is advantageous to sample each subpopulation (stratum) independently. Then simple random sampling or systematic sampling is applied within each stratum.<sup>1</sup>

For example, in the recent 2016 primary, in San Diego County, we noticed a large difference between the results in the various strata – groups of ballots which are received at different times. Particularly in the Early VBM ballots compared with the other groups. In the Early VBM ballots, Hillary Clinton got 64.06% of the vote when compared with Bernie Sanders, but in the Polling place ballots, Later VBM, and Accepted Provisional ballots, she got 44.63%, 50.04%, and 37.46%

1 [https://en.wikipedia.org/wiki/Stratified\\_sampling](https://en.wikipedia.org/wiki/Stratified_sampling)

respectively. Those are very different results in each stratum and the notion that the Early VBM results are sometimes a realistic estimate of the results by media outlets, is false.

We find it unacceptable if the SOS is allowed to exclude any major strata from the scope of the audit. Each strata should be sampled proportionately with its size. This should be stated in the law.

3. We are worried that this law will go into effect without the regulations by the SOS being determined. Such has been the case with the random selection mechanism for the 1% manual tally. The SOS has had the responsibility to determine standards for the random selection mechanism but has done nothing, ever.

We ask that 1) the law not go into effect until the SOS has determined the regulations, 2) there be a mechanism for the public and stake holders to petition the SOS for changes in those regulations, and 3) that regulations by the SOS be set and unchanged at least 90 days prior to any election so citizens can provide proper oversight.

I would rather see the law define the limits for acceptable regulations, not just leave it up to the SOS to decide whatever he/she wants.

4. This version of the bill says that a 5% risk limit is acceptable. That means that 5% of the time (i.e. one out of 20 elections), we may accept a result which is incorrect, and which a full hand count might reverse. That seems like a fairly loose constraint. This should be 1%.
5. Sometimes, we think a full Ballot Image Audit will be better and cheaper. The RLA technique will result in extremely costly ballot-by-ballot sampling when the sample size becomes very large, in the case of a very close election. It appears that there should be a certain threshold where a full manual count will occur when a given race is extremely close and the sample size would otherwise become very large.

So if the cost for pulling a single random ballot is  $C(\text{sample})$  and the cost for doing a full manual count is  $C(\text{full})$ , and if the total number of ballots is  $T$ , then at some  $n \ll T$ ,  $n * C(\text{sample}) > C(\text{full})$ . In other words, pulling ballots one at a time is more costly than doing all the ballots long before the escalation process may indicate that a full manual count is required.

It is acceptable to COPs if the statute allows for (perhaps third-party) verification of the result based on ballot-images coupled with sampling of the correlation between paper ballots and ballot images. In fact, it is our hunch that this will actually be very frequently more cost effective than the ballot-sampled risk limiting audit in those cases when the race is very close. The threshold can be determined in advance where ballot-sampled RLA will be abandoned for a 100% ballot image audit.

6. It is unclear how exactly the ballots are pulled and whether they are replaced into the secure store of ballots, so that if someone were to file a contest of the election, then they could access the stored ballots in their pristine condition, i.e. without any ballots missing due to removal for the RLA process. Or is it the case that the ballot store is missing the ballots pulled for the RLA, and they are separately stored? This is a detail that should be included in the SOS regulations, and it is not mentioned here.

Current regulations states that ballots must be stored by PRECINCT and yet many election districts store them by mixed-precinct BATCH. For example, San Diego and San Bernardino store VBM

ballots by batch. This is likely an oversight in the law.

7. The RLA process must compare the sampled ballot with the digital image of that ballot, because it is relatively easy to modify a ballot (i.e. pencil mark) whereas it is much more difficult to modify ballot images (esp. if they are secured using secure hash digests, as explained below), and if there is a flood or other catastrophe, then the images can be used as a back up to the ballots. If there is a difference between the ballot and the image, and it is an added mark (to result in an over-vote), then the ballot image should be relied upon, and not the potentially modified paper ballot.
8. If used by the election processing equipment, digital ballot images (i.e. “high resolution” full-ballot digital images) must be saved (not deleted), so as to allow the comparison mentioned above, and the potential for a 100% ballot image audit. Such ballot images should be maintained as a permanent record of the election and be available for anyone who wishes to review the election. Such ballot images have already been provided by some counties, such as Dane County, WI and in New York, and are recognized by the Election Assistance Commission as accessible public records.
9. Ballot images should be subjected to an image fidelity test procedure, such as is defined by AIIM TR-34 “Sampling Procedures for Inspection by Attributes of Images in Electronic Image Management (EIM) and Micrographic Systems” (1996). These procedures start with relatively high sampling rate early in the scanning process, and once the quality level is established, then the sampling is done randomly, but less often just to check that the quality is maintained at a high level. The sampling procedures use a set of attributes of the images (such as clarity, contrast, lack of extra lines or artifacts, etc), and do not focus in on the ballot selections or "meaning" of the ballots.
10. Ballot Images should be secured using a two-step hash procedure. 1) create a secure hash message digest (H) of each ballot image and list these in a file of standard format, and 2) create a secure hash message digest of the file containing the list of hash codes, for the lot of ballots being processed (where lot is a precinct, mixed-precinct batch, or portion thereof, processed as a group). These message digests should be published on the election office’s website on a regular basis as ballots are processed, approximately daily. They can be published even as the Early VBM ballots are processed because it is impossible to reconstruct the ballot image from the message digest and yet each will be unique for each ballot image. It must be possible to reproduce the secure hash message digest for each ballot image in the lot and then the digest of digests for the lot. This process will eliminate the possibility that ballot images are modified or added/subtracted from the lot after the hash codes have been published. SThis is similar to the block-chain methodology used in cryptocurrencies (like BitCoin) but without the complexity of those systems based on the need to allow incremental transaction additions and the need to maintain multiple block chains. (See the attached Technical Brief).
11. Election Code should be modified to reflect the fact that
  1. Paper ballots are public records, are not exempted for access by the California Public Records Act (CPRA, Cal Code 6250 et seq) and should be accessible for public review (while remaining under the control of the election officials yet available for review under time, place and manner restrictions), such review may include photographing, scanning and copying. Current law says they are sealed and yet must be stored for 22 months. There is no purpose in keeping ballots for 22 months if they are sealed and no one can review them. The election code predates both the CPRA and the 2004 Constitutional Amendment Article 1 Section 3:

A statute, court rule, or other authority, including those in effect on the effective date of this

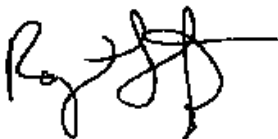
subdivision, shall be broadly construed if it furthers the people's right of access, and narrowly construed if it limits the right of access. A statute, court rule, or other authority adopted after the effective date of this subdivision that limits the right of access shall be adopted with findings demonstrating the interest protected by the limitation and the need for protecting that interest.

2. Digital images of ballots are public records, and should be made available to the public, including the secure hash codes mentioned above to insure that no modifications of the ballot images are possible. Availability of digital ballot images should be explicitly expressed in law.
12. The bill AB-2125 says that the SOS will also determine regulations to "require elections officials to establish appropriate audit boards to conduct the risk-limiting audits." Again, this is too open at this stage. How large will the boards be, how will their members be determined, etc. so they are not determined to be biased toward one party or another.
13. Also the SOS must determine regulations to "ensure the security of the ballots, the selection of ballots to be inspected during each audit, and the rules governing cast vote records and other data involved in risk-limiting audits" and "Establish the calculations and other methods to be used in the audit to determine whether or when the audit of any contest is required to include the examination of more ballots, and to establish calculations and methods to be used in such an escalation, and to determine whether and when the audit of each contest is complete."

There is a great deal going unstated at this phase. I would rather see a draft of the SOS regulations that is being developed in concert with this bill, so it can be discussed as a part of the public process of determining law rather than hoping all will work out later.

14. Specific textual changes to the bill
  1. section 1 (b): Change "cost-effective scientific quality verification" to "cost-effective statistical verification".
  2. 15365: Change "conduct a comprehensive end-to-end verification of software used in the post-election audit process" to "conduct a statistical check of the election results."
  3. 15367 (c): append: "The public will be able to video record and/or broadcast the audit process since there are no voter-identifiable marks on the ballots."

Sincerely,



Raymond Lutz  
National Coordinator, Citizens' Oversight Projects

Citizens Oversight, Inc. is a 501(c)3 Delaware corporation with mission to encourage civic engagement, and focuses in areas with high technical content, such as Election Integrity.

# TECHNICAL BRIEF – PROVIDING AND SECURING DIGITAL BALLOT IMAGES

2018-05-26

Ray Lutz, CitizensOversight.org  
raylutz@citizenoversight.org 619-820-5321

## **INTRODUCTION**

Digital images of paper ballots is used by “next generation” ballot scanner equipment because they can employ much more advanced image processing to determine the voter-intent. Paper ballots can be destroyed by flood or fire, are expensive to store, and can be easily modified by anyone with a pencil. Ballot images, can be inexpensively stored, and once secured, are impossible to modify without detection.

The method for securing the ballot images should be simple and easily reproduced. It should be utilized as soon as practicable after production of the ballot images. We suggest one simple method below, which is similar to the methods used in crypto-currencies, such as BitCoin, but without the additional complexities not needed in this application.

## **IMAGE FILES IN A LOT**

We assume here that ballots are scanned in "work-units" or "LOTS." A LOT can be any convenient group of ballots, perhaps a precinct or batch of vote-by-mail (VBM) ballots.

After the LOT is completed, you will have a set of image files. These files may be simple bit-map format, like .pbm, or some other image file format, such as PDF, TIFF, PNG, etc. There is some valid arguments that the image file should be as simple as possible so there are no hidden crevices where information can be stored as files like PDF, TIFF, PNG, JPG, etc have hidden meta-data which is not immediately apparent. A file format like .PBM is very simple and has no places to hide any information, and once zipped, are still an economical way to store the data.

For example, we will use the ballots published by Dane County, WI, in 2016. Considering “Dunkirk Town Wards 1-6” as the LOT, it has 2,458 images, one for each side of the ballot. A naming convention is used to pair the front and rear images using F and R, as the last letter in the main file name, and the naming convention should also provide the precinct and ballot style. The naming convention used is beyond the scope of this technical brief. Each lot is compressed as a single ZIP archive.

The folder “Dunkirk Town Wards 1-6” looks like this:

```
N0000180000DS01133903640057bb346d664cbF.pbm
N0000180000DS01133903640057bb346d664cbR.pbm
N0000180000DS0113390364006cbd561ff562eF.pbm
N0000180000DS0113390364006cbd561ff562eR.pbm
N0000180000DS0113390364008adffa209c025F.pbm
.
w0000190000DS011339036443abde9210ca4f8F.pbm
w0000190000DS01133903644c2980e95731812F.pbm
w0000190000DS01133903646c6f11a32e2f294F.pbm
w0000190000DS01133903646d01e3d0350a2a2F.pbm
w0000190000DS01133903647bc9989f1491128F.pbm
```

Image data for any lot should be available as a ZIP file on the web site of the election district no later than the day they are scanned (if they are scanned after the election day) or if they were scanned prior to or on election day, then they should be published after election night tabulation is completed.

## **LOT MESSAGE DIGEST FILE**

The first step to securing the images is to create a secure message digest for each ballot image file. To

generate the secure hash message digest file for all files in this folder, the following command can be used. Here, we will use the MD5 secure hash algorithm<sup>2</sup> which is easily available as the program md5sum.exe<sup>3</sup> for windows, and most Linux distributions include it as a standard utility.

This command

```
md5sum *.* > ../LMD_Dunkirk_Town_wards_1-6.txt
```

Creates the file 'LMD\_Dunkirk\_Town\_wards\_1-6.txt' in the parent directory, which contains:

```
907b1311c99d6ae2d5a7d688d1aad39c *N0000180000Ds01133903640057bb346d664cbF.pbm
8a58dfec42e55c81add0135e90d4217b *N0000180000Ds01133903640057bb346d664cbR.pbm
5b65037899b39a9dfa56736f49e21aca *N0000180000Ds0113390364006cbd561ff562eF.pbm
09463082ab83a3a8219d482d34ab3e68 *N0000180000Ds0113390364006cbd561ff562eR.pbm
2a8d5ce74606b276d3f954d1be756a1b *N0000180000Ds0113390364008adffa209c025F.pbm
. . . (snip)
a6de330151abe0e66d483055df595848 *w0000190000Ds011339036443abde9210ca4f8F.pbm
f5680d05dc9a8907dca36670e1719682 *w0000190000Ds01133903644c2980e95731812F.pbm
302c9221be80913c6daf74ca8eac8641 *w0000190000Ds01133903646c6f11a32e2f294F.pbm
f280df7b74759957066f646b9149049e *w0000190000Ds01133903646d01e3d0350a2a2F.pbm
cae435c887c74084dd402c9dfb4f75cb *w0000190000Ds01133903647bc9989f1491128F.pbm
```

Each line of this file provide a secure hash message digest followed by \* and then the file name. We should note that it took about 3 minutes to create all the message digests for 2,458 files on a fairly fast PC, or about 73ms for each file. We will call this the Lot Message Digest (LMD) File, and there should be one per LOT. (The command should be issued so that it lists only the file name without any path.) The process of creating the LMD file should include two workers to reduce any error or malfeasance.

We will note here that the MD5 Secure Hash Digest algorithm, defined in 1991, is not recommended for modern cryptography because there is some remote chance that the digest will be the same for two files that are in fact different, and that it may be able to determine the message from the digest. Because of the nature of this application (the low consequence level if one digest is compromised) it is our opinion that the MD5 algorithm is sufficient and can reduce time costs generating them. But if another (stronger) algorithm is used, it should be expressed as a standard and documented on the website of the election office.

We must realize that here, even the same ballot scanned twice will likely produce different digital images (and thus different digest values), while ballots that are machine generated may generate the same image file and thus the same digest, no matter how strong the algorithm might be, and yet be considered unique. Hand-marked paper ballots will tend to provide enough uniqueness so no two ballots will be digitally identical. Uniqueness can be added, such as an imprinted ballot ID number.

During the canvass period, there should be a separate ZIP file of the folder of the LMD files, for each day that information is released. The file name should have the date that it is completed.

If this message digest is published and others make copies of what the election office has done, then it is impossible to add or alter any of the image files in any of the lots, nor to add and subtract lots, without detection.

### **ELECTION MESSAGE DIGEST FILE**

During the election, lots will be incrementally processed, and one LMD file will be added to a folder which contains all the LMDs for the election so far. After each day, an "Election Message Digest" (EMD) file should be created in a similar manner to the command used above, which has one line for each LMD file.

2 <https://en.wikipedia.org/wiki/MD5>

3 <http://www. etree.org/md5com.html>

That file will provide the Secure Hash Message Digest for each one of the LMD files (which contains, in turn a list of secure message digests and the filename of each image file). In this example, we assumed there are two lots included (so far) in the election, and the EMD file is shown below.

Use this command, where date is filled in with the date of completion.

```
md5sum LMD*.txt >EMD_date.txt
```

Results in this file:

```
ba30f2a4ef65dae434b70d024aa76696 *LMD_Dunkirk_Town_wards_1-6.txt  
1b431d03d49e30129214e54a3a9177dc *LMD-Dunn_Town_wards_1-7.txt
```

After approximately each day (including any days of scanning prior to the election for early voting) the election district should publish a new EMD\_date.txt file. Compared with the prior day, each line in this file will not change as the election is completed, lines are added as each lot is processed. Each EMD daily file should be published on the website of the election district (and not published by updating a single file).

It must be emphasized that the EMD files must NOT be coupled or embedded with cast vote record (CVR) data and should be published separately, and prior to any final disclosure of CVR records.

### **OVERSIGHT PROCEDURES**

Any group providing oversight -- including the Secretary of State -- should download the files from each election district each day. They should check that the EMD file provides the same message digest for each LMD entry compared with files for earlier days. After the image files are available, then oversight groups can check that the image data produces the message digest in each LMD file provided. This will eliminate any risk that files can be added, modified or lots changed.

The Secretary of State should gather up all the MD data from each jurisdiction.

With availability of ballot images, any oversight group can determine the results of the election.

### **COMPARISON WITH OFFICIAL RESULTS**

Election officials should prepare a ballot-by-ballot Cast Vote Record CVR with link to the ballot image file, preferably including the same message digest which was provided in the LMD file. Any oversight group that wishes to challenge the results can compare their CVR data with that published by the election district, and provide any specific challenges to official canvass on a ballot-by-ballot basis.

The election office should also provide ballot-styles data, including how the ballot style can be determined by either the image file name or other embedded information, and how the paper ballot can be accessed from secured storage.

More information: <http://citizenoversight.org>