Uniform Audit Report Format

Ray Lutz, Citizens Oversight -- raylutz@citizensoversight.org

V0.1 -- 2019-04-29
V0.2 -- 2019-12-06 -- Ported to GoogleDocs
V0.3 -- 2019-12-10 As submitted to CA RLA Regs Proceeding

1.0 Introduction

Election auditing is commonly done as self-audits by the same election officials who are responsible for certifying the election. In fact, it is preferred that these audits be done prior to certification so the results of the audit can affect the outcome, if something questionable is detected. Of course, if rigorous audits are in place, it is much less likely that anyone will even attempt to hack the election.

But for these self-audits to be effective, they rely on transparency and also on rigorous scrutiny by the public. Is is essential, therefore, that a common audit report format be utilized so that oversight groups can more easily provide oversight by those election officials with a minimum of overhead.

2.0 Goals

The following goals were created for this data format:

1. The report format should be human readable and should provide the official written report of the audit so there is no chance of transcription errors.
2. The report format should be easy for machines to parse.
3. The report should be easy to generate and publish.
4. The report format shall be easy to prepare using standard computer tools, most likely using any popular spreadsheet program. However, more user friendly front-ends are also feasible.
5. The report should provide concise results on page one with underlying results on subsequent pages.
6. The report format should support various audit methods and any risk calculations.
7. The report format should contain key data items and links to associated files so that an outside observer can reproduce the audit to allow independent oversight.
8. The report format is primarily focused on Ballot Comparison, Ballot Polling, and Batch Comparison audits.

9. The report should be suitable for county or statewide application.
10. The audit report format should encourage an auditing approach which discourages "innocent fixup" during the audit process. For example, the format should encourage the values actually read from the ballots to be recorded independently and without reference to the values in the cast-vote record (CVR). Once the voter intent is inspected from the ballots, perhaps from two independent teams, the results from the manual inspection of the ballots can be published and frozen so they are not innocently changed. Only then are actual votes compared with the cast-vote record. This is in contrast with the approach of directly comparing each ballot to each CVR and recording whether there was a discrepancy or not. (This is a key difference between the Colorado audit report which directly recorded discrepancies rather than actual values.
11. The final report should be complete so it can be directly published as the official human-readable audit report. However, it must also be publishable when it is in incomplete form so it is not possible for fraudsters or compromised officials to change earlier information after the fact, that is, to "fix up" the report so it covers up an audit that otherwise would have reported significant discrepancies.

    One way to solve this is to secure and publish incomplete versions of the report at each logical step of the audit process. Another option might be to somehow use hash codes to treat the incremental report like a block-chain. The final resolution of this design issue is still open.

    The use of standard tally sheets which are published and secured will help defray fears that the report can be manipulated.
12. The report format will likely include multiple components. It is not required to be a single file, but some of the components may be separate sheets within the same spreadsheet file. The report will refer to other output and intermediate files, such as Cast-Vote Records file (CVRs) and other reports produced by election management software. The format of these files is generally out of scope of this audit report standard, but some of the intermediate files which are needed to perform oversight may be defined here. This standard may recommend, limit the formats to be used or profile the use of those files to recommended subsets.
13. A separate component of the report to be standardized is a "landing" URL object which will be constant for any given election district, and, preferably, in only several common locations relative to the election district website. This landing URL will provide the location of the audit report and other commonly provided election reports. The data item at this URL should be in RSS 2.0 format (XML) or perhaps JSON version of that standard that has been recently proposed.

14. The report format will be ordered according to the expected audit procedure to allow use of the report format to guide implementation of audits.
15. The format should be incrementally immutable such that, once fields are completed, they need not change if the file is revised.
16. The format should include fields that explain each step so that it can be used as a checklist as audit procedures are performed by workers.
17. Line items that represent tasks or data that are expected to be available for a given audit protocol will encourage audit workers to recognize whether those items are being provided, and if not, provide that information in the future.
18. URLs provided in format fields shall point to publicly accessible resources and will not require security credentials to access.
19. Details for launching helper applications will be supported for each line item in a form that may allow more user-friendly step-by-step software to be layered on top of the format in a generic fashion such that said software is not cognizant of what the step actually does, and, therefore, is feasible to limit the risk if that the front-end becomes an additional risk factor.
20. The format will allow a more comprehensive evaluation of confidence based on the manner in which the audit is performed and data provided for inspection by the public.

2.1 Use Cases

The following use cases are identified to be supported by the audit report format:

1. Step-by-step guide for workers implementing the audit, by referring to the "empty" template report which has instructions for each step but no values filled in.
2. More user-friendly front-end can be then developed which uses the information provided in the template report so as to provide the instructions in step-by-step dialog boxes and offer more checking of input values to reduce human error in completing the report, but which does not introduce custom software with additional risks.
3. Report can be processed by a software evaluator that can provide feedback to those completing the audit if values are of the wrong type or inconsistent.
4. Report should be posted to a public location. As the audit is processed and incrementally improved, each version of the report shall remain available, which may be part of the posting facility.
5. Report format shall support automated oversight by organizations that wish to review the data and bounce it off their own consistency checks.

6. Report shall support self scoring of overall risk and scores for transparency and completeness.

3.0 Existing Report Formats

There is no standard audit report format at this time. However, a number of states conduct audits and there are some ad-hoc formats that have been used. Colorado conducts risk-limiting audits on a statewide basis for a limited number of races, typically one state-wide race and one county-wide race for each county. California and Florida have conducted batch comparison audits. California samples 1% of the precincts and vote-by-mail ballots, and must sample at least one precinct for all races. Florida chooses just one race randomly, but no audit is performed if a machine recount is mandated. A number of risk-limiting audit pilots have been conducted, such as in Orange County, CA and Rhode Island. Orange County included only ballot-polling while Rhode Island investigated ballot-polling, ballot-comparison, and batch-comparison audits.

As the fields are described below, comparison with these other formats will be included if we have the information. The Colorado audit report is the most similar to the report format proposed here as their final report was in a spreadsheet format. Orange County provided only a narrative report.

4 General Format

4.1 XLSX File format

There is precedent to publish the audit reports as a spreadsheet, i.e. compatible with popular spreadsheet programs such as Excel, LibreOffice Calc, etc. These can also be expressed as a set of CSV (comma separated values) files, one per sheet of the workbooks provided in those programs. Because of its popularity, the fact that it can accommodate multiple sheets, and deals gracefully with embedded newlines, the Excel format will be adopted with UTF-8 encoding, as it is now the most accepted character encoding.

4.2 Sections

The format includes the following sections:

1. **Header** -- The header section is in parameter-value format and provides format metadata or district-wide data. This section also provides links to other files, such as Cast-Vote Record file (CVR), ballot manifest, and other source files with their formats and hash codes.
2. **Contests, official and audit results** -- This section is formatted to allow one-line per ballot option in each audited contest, including the audit results and calculated risk.

3. **Data entry sheets** -- This data is transcribed from tally sheets which collect the data and are separately secured. Subsequent sheets of the workbook provide actual data entered by the auditing team.

An example file that meets this format definition is at this link: https://docs.google.com/spreadsheets/d/12qF57LoiWTou_j4HZ_QqYaOz-AlsD6jcks71YItkhKw/edit?usp=sharing . This report reflects the results of the Orange County, CA pilot audit in 2018.

5. Summary Page

The first sheet in the report file is the Summary page. This page provides summary information that applies to the entire election in this district. The information is provided in the following columns:

1. Parameter - is the name of the parameter and must be exactly as shown
2. Value - is typically a string, filename, or URL.
3. Format - may provide a format descriptor
4. Hash - provides the SHA256 (or other) hash digest of the entire file referenced.

The first line should contain the column names Parameter, Value, Format, Hash. Here, we also include the Description column to allow definition of the content of the record line.

Parameter	**Value**	**Format**	**Hash**	**Description**
Format	Audit Report	MM.mm		The string self identifies this report. Format column is the version number of the Audit Report format. Current value is 00.01
Format Definition	(URL)	(this file)		This entry is to be provided once this standard is completed. This might be in a schema syntax.
Generation DateTime	Use spreadsheet format: 1999-12-31 13:37:46			RFC-3339 where "T" character may be a space. CO audit report used "11/21/2018 02:25 PM" format without timezone. Spreadsheet date can be accessed with control-semicolon or shift-control-semicolon
Completion Status	Final | Partial Phase n			String Enumeration, describes the completion status of the audit report. "Final" is the final report with all information completed. "Partial Phase 1"

				includes values required for phase 1 completion.
District Type	county | state | municipality | national			String Enumeration, using values from NIST CDF ReportingUnitType
District	CountyName, ST|ST			Name of the district providing this audit report.
Election Date	YYYY-MM-DD			
Election Type	general partisan-primary-closed partisan-primary-open primary runoff special other			Same definition as in NIST CDF ElectionType
Posting Site	URL	Sharefile Linkpage		Link to a posting service like Sharefile.com where files can be found. Such posting services should offer trusted timestamps so no date can be forged. If provided, subsequent file reference can be file names that exist on the URL provided (Linkpage)
Narrative Report	URL | Filename	PDF	SHA-256 Hash of Entire file.	(optional) any additional narrative audit report.
CVR	URL | Filename	ES&S Dominion NIST_CVR	[HashType:]Hashcode[;CK] Standard HashType is SHA256. Resulting hash will be either raw hex digit string for SHA256 or preferred, the HashType, followed by the HashCode as hex digits, followed by hex checksum digits for the byte to be added to the byte sum of the hash to sum to 0.	Cast-vote record files should be limited to no more than 100,000 lines and split into separate files as needed to allow the files to be opened with conventional computers. Multiple lines can exist to reference multiple files. Simple file name implies it can be found at the posting service site.
Manifest	URL | Filename	CSV|XLSX	SHA256	The manifest file and its format and hash. Format: fields="BallotID,Batch,Sheet"

Audit Type	ballot_comparison ballot_polling batch_comparison ballot_image_audit			
Risk Limit	5.0%			percent, written with percent sign for readability.
Seed Selection Method	URL to beacon or method description	public dice roll | random beacon		
Seed Selection Notice URL	URL to PDF	PDF | HTML	SHA256	This document should be the official announcement of the seed selection meeting.
Seed Selection Meeting Date-Time	2018-11-16 10:00+07:00			RFC-3339 or Spreadsheet format (control-semicolon)
Seed Selection Video URL	URL to video (such as YouTube)			
Random Seed Posting URL	URL to PDF or filename	PDF		Official posting of the random seed
Random Seed	example: 64496045949432238293			String of digits. Must allow leading zeros to be preserved. Typ. at least 20 digits.
Name of PRNG	Sampler (Rivest)			Must provide the algorithm name rather than "on the website of ..."
Number of Sheets per ballot	integer			Integer
Total Ballot Cards in Manifest	integer			CO uses Ballot Cards
Total Ballot Cards in CVR	integer			CO uses Ballot Cards
Upload Status URL	URL to CSV or XLSX	CSV or XLSX		This file provides a log of uploaded manifest and CVR files.
Upload Status Format				
Ballot Images	filename.zip			Ballot image archives should be in zip format and limited to no more than about 5GB
Ballot Image Source Hashes	filename.zip			This file contains the ballot image name and the original source image hash. This hash should be created by the source device of only the core image of each page, and can be checked against the core

				image "inside" any file wrapper such as PDF, TIFF, PNG, etc.
EIF	filename.xlsx			This is the election information file which provides the list of all contest names and options as found in the CVR, on the Ballot itself, and on any BMD ballot summaries, if applicable, indexed to the official contest and option names. This also provides the full text description as found on the ballot, and in any supported languages other than English.
Style Description File				For each style used in the election, this file provides the set of contests found on the ballot for that style in the order they are printed.
Initial Sample	integer			
Picklist	URL or filename(s)		Hash	
Tally Sheets	URL or filename(s)		Hash	

6. Results Sheet

The results sheet provides both the reported results and the results of the audit, and provides the calculated risk limit. The columns of this page will vary depending on the type of audit.