Data Export

Colorado Risk Limiting Audit Tool

2017

Table of Contents

FREE & FAIR

Separate from the RLA application server and client software that supports the Department of State and the Counties in carrying out the Risk Limiting Audit, there is a command, called `rla_export`, allowing export of data from the central server and the underlying database.

The command is part of a python package, whose technical description can be found in a `README.rst` file in the python site-packages directory tree wherever the package is installed. The most current online version (which may or may not match the version you have installed) is available in the public code repository.

The `README.rst` file gives instructions for installing the python package, and describes various run-time options. For a catalog of the exports produced by the command, see below.

Minimum Data Required to Allow Public to Reproduce Audit Calculations

To allow independent verification of the RLA, the Colorado Department of State must provide to the following to the public :

- all CVR files and their hashes
- all ballot manifest files and their hashes
- the list of contests selected for audit, and which if any have been designated for hand count
- opportunity to observe the random seed selection
- for each County, the random sequence of ballot cards determined by the random seed and the pseudo-random number generator
- opportunity to observe the activities of the County Audit Boards
- announced tabulated results for contests selected for audit
- the risk limit
- the error inflation factor
- the tabulation of results and counts of ballot cards used to calculate the diluted margin (including the number of winners for each contest selected for audit)
- for each County and each round of the audit, the list of ballot cards assigned to the Audit Board for review
- for each contest selected for audit, and for each cast vote record that contains the given contest and has been presented to the Audit Board for verification, the RLA system's record of the Audit Board's review of the physical ballot for that contest

Exports

The data files in this section are generated based on `sql` query files. These are always produced in two formats: json and csv. The basename of each resulting file is the same as

the basename of the query file. Thus, given the query file `seed.sql`, the files `seed.json` and `seed.csv` will be produced.

Database Exports in.csv and .json format

The `rla_export` command exports many of the files necessary for independent verification of the RLA, whether by candidates, parties, other organizations. Specific exports are detailed below.

m_tabulate.sql

Field	Type	Meaning
county_name	Text String	Name of County
contest_name	Text String	Name of contest
choice	Text String	Name of candidate or for a ballot question "Yes" or "No"
votes	Integer	Number of votes recorded for the given choice in the given contest in the given County

m_selected_contest_audit_details_by_cvr

For each contest under audit, and for each cast vote record that contains the given contest and has been presented to the Audit Board for verification, the RLA system's record of the Audit Board's review of the physical ballot for that contest.

Note that the number of discrepancies each cast vote record contributes to the risk level calculation depends not only on the discrepancies found between the cast vote record and the Audit Board interpretation, but also on the number of times that cast vote record counts in the random sequence.

Field	Type	Meaning
county_name	Text String	Name of County
contest_name	Text String	Name of contest
imprinted_id	Text String	Combination of scanner number, batch number and position within batch that uniquely identifies the ballot card and may be imprinted on the card when the ballot is scanned
ballot_type	Text String	BallotType from Dominion CVR export file, a code for the set of contests that should be present on the physical ballot card. Often

		known as *ballot style*.
duplicates_in_random_sequence	Integer	Number of times that a discrepancy between the cast vote record with the given imprinted id and the audit board interpretation has been counted toward the risk level
choices_per_voting_computer	List of Text Strings	List of voter choices in the given contest on the given ballot card, as interpreted by the vote-tabulation computer system (note: overvotes recorded as blank votes)
choices_per_audit_board	List of Text Strings	List of voter choices in the given contest on the given ballot card, as interpreted by the Audit Board (note: overvotes recorded as a too-long list of choices)
consensus	YES/NO	YES if the Audit Board came to consensus on the interpretation of the given ballot card; NO if not; blank if the card has not been reviewed by the Audit Board.
audit_board_comment	Text String	Text of comment entered by Audit Board about the given contest on the given ballot card, or indication that the ballot was not found.
timestamp	Timestamp	Date and time of Audit Board's submission of their interpretation to the RLA Tool
cvr_id	Integer	Internal database id for the cast vote record

m_selected_contest_static

List of contests selected to drive the audit, with information about the contest that doesn't change during the audit, namely the reason for the audit, the number of winners allowed in the contest, the tabulated winners of the contest, the numbers of ballot cards recorded as cast in the county (total number as well as the number containing the given contest) and the value of the error inflation factor (gamma).

Field	Type	Meaning
county_name	Text String	Name of County
contest_name	Text	Name of contest

	String	
audit_reason	Text String	Reason for audit (STATE_WIDE_CONTEST and COUNTY_WIDE_CONTEST refer to the types of contests that must be chosen to drive the audit, per Rule 25.2.2(i))
votes_allowed	Integer	Maximum number of choices that can be recorded on one ballot in the given contest
winners_allowed	Integer	Number of winners allowed for the contest (required to calculate diluted margin)
winners	List of Text Strings	List of all winners of the given contest in the given County. (Note that for multi-county contests this list includes the highest vote-getters within the County, even if these were not the winners across all Counties.)
min_margin	Integer	The smallest margin between any winner and any loser
county_ballot_card_count	Integer	The number of ballot cards recorded in the given County in the election (including cards that do not contain the contest in question)
contest_ballot_card_count	Integer	The number of ballot cards recorded in the given County that contain the contest in question
gamma	Number	Error inflation factor defined in Stark's paper, Super-simple simultaneous single-ballot risk-limiting audits, which is cited in Lindeman and Stark's paper, A Gentle Introduction to Risk Limiting Audits, which is cited in Rule 25.2.2(j))

m_selected_contest_dynamic

List of contests selected to drive the audit, with current status. Which contests (if any) have been selected for hand count? How many discrepancies of each type are there?

Field	Type	Meaning
county_name	Text String	Name of County
contest_name	Text String	Name of contest
current_audit_type	Text String	COMPARISON, HAND_COUNT, NOT_AUDITABLE or NONE. Note that NOT_AUDITABLE means the contest can't drive an audit, but it still can be audited opportunisticly.
random_audit_status	Text String	NOT_STARTED, NOT_AUDITABLE, IN_PROGRESS or ENDED. Because declaring a hand count ends the computerized portion of the audit, a contest that is being hand-counted will have the value ENDED in this field.
one_vote_over_count	Integer	The number of ballot cards in the random sequence so far (with duplicates) on which there is a one-vote overstatement (per Lindeman & Stark's A Gentle Introduction to Risk Limiting Audits).
one_vote_under_count	Integer	The number of ballot cards in the random sequence so far (with duplicates) on which there is a one-vote understatement (per Lindeman & Stark's A Gentle Introduction to Risk Limiting Audits).
two_vote_over_count	Integer	The number of ballot cards in the random sequence so far (with duplicates) on which there is a two-vote overstatement (per Lindeman & Stark's A Gentle Introduction to Risk Limiting Audits).
two_vote_under_count	Integer	The number of ballot cards in the random sequence so far (with duplicates) on which there is a two-vote understatement (per

	Lindeman & Stark's A Gentle Introduction to Risk Limiting Audits).

m_cvr_hash

Hashes of CVR files

Field	Type	Meaning
county_name	Text String	Name of County
cvr_export_hash	Text String	Hash value entered by the given county after uploading the cast vote record file to be used in the audit

m_manifest_hash

Hashes of ballot manifest files

Field	Type	Meaning
county_name	Text String	Name of County
ballot_manifest_hash	Text String	Hash value entered by the given county after uploading the ballot manifest file

to be used in the audit

all_contest_static

List of all contests, with information about the contest that doesn't change during the audit, namely the reason for the audit, the number of winners allowed in the contest, the tabulated winners of the contest, the numbers of ballot cards recorded as cast in the county (total number as well as the number containing the given contest) and the value of the error inflation factor (gamma).

Field	Type	Meaning
county_name	Text String	Name of County
contest_name	Text String	Name of contest
audit_reason	Text String	Reason for audit (STATE_WIDE_CONTEST and COUNTY_WIDE_CONTEST refer to the types of contests that must be

		chosen to drive the audit, per Rule 25.2.2(i); other reasons from the Rule include CLOSE_CONTEST, TIED_CONTEST, GEOGRAPHICAL_SCOPE, CONCERN_REGARDING_ACCURACY, and COUNTY_CLERK_ABILITY; the audits of other contests on the ballot are OPPORTUNISTIC_BENEFITS)
votes_allowed	Integer	Maximum number of choices that can be recorded on one ballot in the given contest
winners_allowed	Integer	Number of winners allowed for the contest (required to calculate diluted margin)
winners	List of Text Strings	List of all winners of the given contest in the given County. (Note that for multi-county contests this list includes the highest vote-getters within the County, even if these were not the winners across all Counties.)
min_margin	Integer	The smallest margin between any winner and any loser
county_ballot_card_count	Integer	The number of ballot cards recorded in the given County in the election (including cards that do not contain the contest in question)
contest_ballot_card_count	Integer	The number of ballot cards recorded in the given County that contain the contest in question
gamma	Number	Error inflation factor defined in Stark's paper, Super-simple simultaneous single-ballot risk-limiting audits, which is cited in Lindeman and Stark's paper, A Gentle Introduction to Risk Limiting Audits, which is cited in Rule 25.2.2(j))

all_contest_dynamic

List of contests with current status. Which contests has the Secretary selected for audit? Which contests (if any) has the Secretary selected for hand count? How many discrepancies of each type have been found so far?

Field	Type	Meaning
county_name	Text String	Name of County
contest_name	Text String	Name of contest
current_audit_type	Text String	Comparison audit, ballot polling audit or hand count
computerized_audit_status	Text String	Not started, in progress, risk limit achieved, or ended. Because declaring a hand count ends the computerized portion of the audit, a contest that is being hand-counted will have the value "ended" in this field.
one_vote_over_count	Integer	The number of ballot cards in the random sequence so far (with duplicates) on which there is a one-vote overstatement (per Lindeman & Stark's A Gentle Introduction to Risk Limiting Audits).
one_vote_under_count	Integer	The number of ballot cards in the random sequence so far (with duplicates) on which there is a one-vote understatement (per Lindeman & Stark's A Gentle Introduction to Risk Limiting Audits).
two_vote_over_count	Integer	The number of ballot cards in the random sequence so far (with duplicates) on which there is a two-vote overstatement (per Lindeman & Stark's A Gentle Introduction to Risk Limiting Audits).
two_vote_under_count	Integer	The number of ballot cards in the random sequence so far (with duplicates) on which there is a

	two-vote understatement (per Lindeman & Stark's A Gentle Introduction to Risk Limiting Audits).

all_contest_audit_details_by_cvr

For each contest and for each cast vote record that contains the given contest and has been presented to the Audit Board for verification, the RLA system's record of the Audit Board's review of the physical ballot for that contest.

Note that the number of discrepancies each cast vote record contributes to the risk level calculation depends not only on the discrepancies found between the cast vote record and the Audit Board interpretation, but also on the number of times that cast vote record counts in the random sequence.

Field	Type	Meaning
county_name	Text String	Name of County
contest_name	Text String	Name of contest
imprinted_id	Text String	Combination of scanner number, batch number and position within batch that uniquely identifies the ballot card and may be imprinted on the card when the ballot is scanned
ballot_type	Text String	BallotType from Dominion CVR export file, a code for the set of contests that should be present on the physical ballot card. Often known as *ballot style*.
duplicates_in_random_sequence	Integer	Number of times that a discrepancy between the cast vote record with the given imprinted id and the audit board interpretation has been counted toward the risk level
choices_per_voting_computer	List of Text Strings	List of voter choices in the given contest on the given ballot card, as interpreted by the vote-tabulation computer system (note: overvotes recorded as blank votes)
choices_per_audit_board	List of Text Strings	List of voter choices in the given contest on the given ballot card, as interpreted by the Audit Board (note: overvotes recorded as a

		too-long list of choices)
consensus	YES/NO	YES if the Audit Board came to consensus on the interpretation of the given ballot card; NO if not; blank if the card has not been reviewed by the Audit Board.
audit_board_comment	Text String	Text of comment entered by Audit Board about the given contest on the given ballot card, or indication that the ballot was not found.
timestamp	Timestamp	Date and time of Audit Board's submission of their interpretation to the RLA Tool
cvr_id	Integer	Internal database id for the cast vote record

auditboards

Field	Type	Meaning
county_name	Text String	Name of County
member	Text String	Name of audit board member
sign_in_time	Timestamp	Beginning of an audit board member's RLA Tool session
sign_out_time	Timestamp	End of the given session for the given audit board member

batch_count_comparison

Field	Type	Meaning
county_name	Text String	Name of County
scanner_id	Integer	the identification number of a scanner used to create the cast vote record from the physical ballot card
batch_id	Integer	The identification number of a batch of ballot cards scanned by the given scanner
count_per_manifest	Integer	The number of ballot cards in the given batch on the given scanner,

according to the ballot manifest file uploaded by the County count_per_cvr_file | Integer | The number of ballot cards in the given batch on the given scanner, according to the cast-vote-record file export from the voting computer, uploaded by the County difference |

Integer | The difference between the two counts, which will be zero for a correctly tabulated election. If positive, there are ballots listed in the manifest that are not found in the CVR file; if negative, there are ballots in the CVR file that are not listed in the manifest.

prefix_length

Field	Type	Meaning
county_name	Text String	Name of County
audited_prefix_length	Integer	Length of the longest prefix of the random sequence of cvr selections containing only cvrs that have been audited

seed

Field	Type	Meaning
seed	20-Digit String	the random seed for the pseudo-random number generator

upload_status

Field	Type	Meaning
county_name	Text String	Name of County
filename	Text String	Name of file
hash_status	Text String	VERIFIED, MISMATCH, or NOT_CHECKED
approx_count	Integer	Approximate number of lines in the file
size	Integer	Size of file in bytes
status	Text String	IMPORTED_AS_BALLOT_MANIFEST, IMPORTED_AS_CVR_EXPORT or NOT_IMPORTED
timestamp	Timestamp	Date and time of the most recent update to the upload status of the given file

Reports in .xlsx Format

Some files are exported from the application server in .xlsx format.

County Audit Reports

There is a separate report (in .xlsx format) for each County. Within each County's report there is a separate spreadsheet for each round of the audit containing the list of ballot cards

assigned to the County Audit Board for that Round. For each ballot card in the list the spreadsheet indicates whether it has been reviewed, whether any discrepancies were found on the card and whether the Audit Board disagreed on the interpretation of the card. There is a summary page with a variety of audit information, and an affirmation page (which will be blank).

State Audit Report

Within this report (in .xlsx format) there is a separate spreadsheet for each County containing the information from that County's round spreadsheets. The summary spreadsheet contains a variety of audit information, both general and county-specific.

Lists in .csv Format

Some files are exported from the application server in .csv format.

County Ballot Card List by Round

Field	Type	Meaning
scanner_id	Integer	TabulatorNum from Dominion CVR export file, identifying the tabulator used to read the physical ballot card
batch_id	Integer	BatchId from Dominion CVR export file, identifying the batch of physical ballot cards in which the card was scanned
record_id	Integer	RecordId from Dominion CVR export file, indicating the position of the card in its batch of physical ballot cards
imprinted_id	Text String	combination of scanner, batch and record ids that uniquely identifies the ballot card and may be imprinted on the card
ballot_type	Text String	BallotType from Dominion CVR export file, a code for the set of contests that should be present on the physical ballot card. Also known as *ballot style*.
storage_location	Text String	The physical location of the ballot
cvr_number	Integer	The index of the given cast vote record in the CVR file, starting at 1, used to associate lines of the

		CVR file to numbers generated by the pseudo-random number generator
audited	Yes/No	Yes if the ballot card has been reviewed by the Audit Board; otherwise No.

County Random Sequence

Field	Type	Meaning
county_name	Text String	Name of County
round_number	Integer	Round of the audit
random_sequence_index	Integer	The position in the random sequence for the given County
scanner_id	Integer	TabulatorNum from Dominion CVR export file, identifying the tabulator used to read the physical ballot card
batch_id	Integer	BatchId from Dominion CVR export file, identifying the batch of physical ballot cards in which the card was scanned
record_id	Integer	RecordId from Dominion CVR export file, indicating the position of the card in its batch of physical ballot cards
imprinted_id	Text String	combination of scanner, batch and record ids that uniquely identifies the ballot card and may be imprinted on the card
ballot_type	Text String	BallotType from Dominion CVR export file, a code for the set of contests that should be present on the physical ballot card. Also known as *ballot style*.

Technical Notes

List Specifications

Each list of integers are encoded as a json string, i.e., a string enclosed in square brackets ([]) with numbers delimited by commas (,)

Ballots vs. Ballot Cards

When a ballot extends across more than one piece of paper (a "card"), each card is tabulated independently. In Counties which have any multi-card ballots, the ballot card counts provided will be greater than the turnout figures reported elsewhere. For example, in November 2017 the County of Denver had a two-card ballot.