NIST Grant/Contractor Report
NIST GCR 22-034

Recommendations for Voting System Interoperability

John Dziurłaj

This publication is available free of charge from:
https://doi.org/10.6028/NIST.GCR.22-034

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE

NIST Grant/Contractor Report
NIST GCR 22-034

Recommendations for Voting System Interoperability

John Dziurłaj
The Turnout

This publication is available free of charge from:
https://doi.org/10.6028/NIST.GCR.22-034

August 2022

U.S. Department of Commerce
Gina M. Raimondo, Secretary

National Institute of Standards and Technology
Laurie E. Locascio, NIST Director and Under Secretary of Commerce for Standards and Technology

NIST GCR 22-034
August 2022

This publication was produced as part of contract SB1341-15-CQ-0010/1333ND20FN770201 with the National Institute of Standards and Technology. The contents of this publication do not necessarily reflect the views or policies of the National Institute of Standards and Technology or the US Government.

NIST Technical Series Policies

How to Cite this NIST Technical Series Publication

Dziurłaj J (2022) Recommendations for Voting System Interoperability. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Grant/Contractor Report (GCR) NIST GCR 22-034. https://doi.org/10.6028/NIST.GCR.22-034

Abstract

There is currently no common data format for ballot styles and associated use-cases. This white paper investigates this and other potential interoperability gaps within the voting system in a holistic manner, by considering previous voting interoperability work and how it fits within thirteen key functions of a voting system. Gap analysis is performed by identifying data flows that cross functions, functions that might cross component boundaries and thus may significantly benefit from common data interfaces. Each data flow identified as significant for interoperability is treated in detail, describing the potential benefits of supporting interoperable interfaces in voting system components and a roadmap to their development. The gap analysis shows lack of support for data flows that involve ballot data, particularly ballot styles. These gaps preclude effective componentization of voting system devices such as ballot marking devices, election management systems and scanners. Identified solutions include standardization of ballot components, such as barcodes, and development of a ballot styles common data format.

Key words

Ballot printing systems; Ballot scanners; Common data format; Electronic poll books; Election technology; Interoperability; Voting systems.

Table of Contents

List of Tables

List of Figures

1. Interoperability inside the Voting System

Past work on common data formats focused on peripheral input and outputs between the voting system and its environment. Lack of standardization of these interfaces has been a perennial pain point for election officials. In response, NIST developed several common data formats that tackled interfaces to support election results reporting, election event logging, and cast vote records.

Most of these specifications support communication between the voting system and non-voting system components. The specifications have only addressed interoperability between components within the voting system in limited areas, such as cast vote records and election event logs.

In contrast to past efforts, this document focuses on interoperability requirements *within* the voting system. We accomplish this by examining the functions that a voting system must perform and the data required to support those functions. In this paper we identify significant data flows and analyze the feasibility of supporting interoperability between components performing major voting functions. Finally, we provide suggestions for future work to improve interoperability in the identified areas.

1.1. Scope

This document considers the interoperability requirements of voting system devices that could potentially meet VVSG 2.0 requirements[2]. It also discusses electronic poll books, whether or not they are part of a certified voting system, as they usually perform the *determine ballot* function in modern polling location configurations. An auditing use case was suggested and discussed in the research group. However, since this particular effort does not consider components that fall outside the defined functions of the voting system, auditing did not fall within scope.

1.2. Identifying Interoperability Requirements

The interoperability requirements within the voting system are identified through structured analysis. Specifically, a data flow diagram is constructed, using Gane-Sarson notation[1], with each function of a voting system (see below) listed as a "*process*". The data required for each function is specified using *data flows* – arrows pointing from the source of the data to its the target. Data can flow between *processes*, *info stores*, and *external agents*.

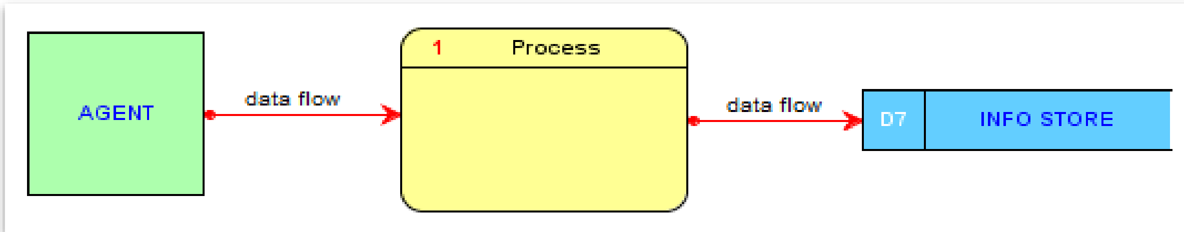


Figure 1 - Shapes used by data flow diagrams

Required data inputs and outputs for each function will vary based on the ballot marking method used. For example, a voter using a ballot marking device (BMD) will mark their ballot using electronic screens rather than a pen as when using a paper ballot. To better delineate these flows, we have colored them differently based on whether the scenario is hand marked paper ballot (HMPB, in blue) or BMD (in green).

Next, the data flow diagram is overlaid with the voting system components that could provide each function (see Section 3).

This brings the interoperability requirements into full view: interoperability is indicated when there is a flow that crosses a component boundary.

The specific set of components supporting each function will vary from scenario to scenario. Several common scenarios are elaborated, including a BMD scenario and a HMPB scenario.

The data flows that cross a component boundary and that are identified as significant for interoperability are analyzed for current CDF coverage and considerations for enabling interoperability.

Finally, a set of future steps are outlined.

2. Functions of a Voting System

The VVSG [2] defines a voting system as:
Equipment (including hardware, firmware, and software), materials, and documentation used to enact the following functions of an election:

1. *define elections and ballot styles,*
2. *configure voting equipment,*
3. *identify and validate voting equipment configurations,*
4. *perform logic and accuracy tests,*
5. *activate ballots for voters,*
6. *record votes cast by voters,*
7. *count votes,*
8. *label ballots needing special treatment,*
9. *generate reports,*
10. *export election data including election results,*
11. *archive election data, and*
12. *produce records in support of audits.*

From this definition, thirteen functions of a voting system have been identified:

No.	Use-case	Description
F1	Input Ballot Data	Input the contests and candidates that appear on the ballot, along with associated data
F2	Layout ballot	Layout the data necessary for constructing a ballot
F3	Generate ballots	Print paper ballots with ballot style and activation information (if applicable)
F4	Determine ballot	Determine which ballot styles the voter is eligible, and handle their issuance
F5	Present ballot	Present the ballot to the voter in the format requested
F6	Capture vote selections	Capture the voter's selections as they are made
F7	Interpret vote selections	Apply business rules according to voting method
F8	Present vote selections	Present vote selections or second chance criteria for user to confirm
F9	Cast to paper	Store to a paper record
F10	Retrieve vote selections	Read paper ballot
F11	Store selections electronically	Store votes to device memory
F12	Tabulate vote selections	Tally the ballot selections
F13	Present results	Provide up to date election results

Voting system components can be defined in terms of the functions they perform. While it may be possible to mix functions in arbitrary ways to produce many unique devices, it is likely that real world devices will compose functions similarly. For example, a scanner might:

- *Retrieve vote selections,*
- *Interpret vote selections,*
- *Present vote selections*, and
- *Store selections electronically*

Almost all functions can appear independently of others in a component.
(Note that the numbering of the functions follows a rough order that these functions may occur. It does not imply that these functions always occur in that order.)
We have constructed a data flow diagram showing the thirteen functions and data dependencies between them in Figure 2.

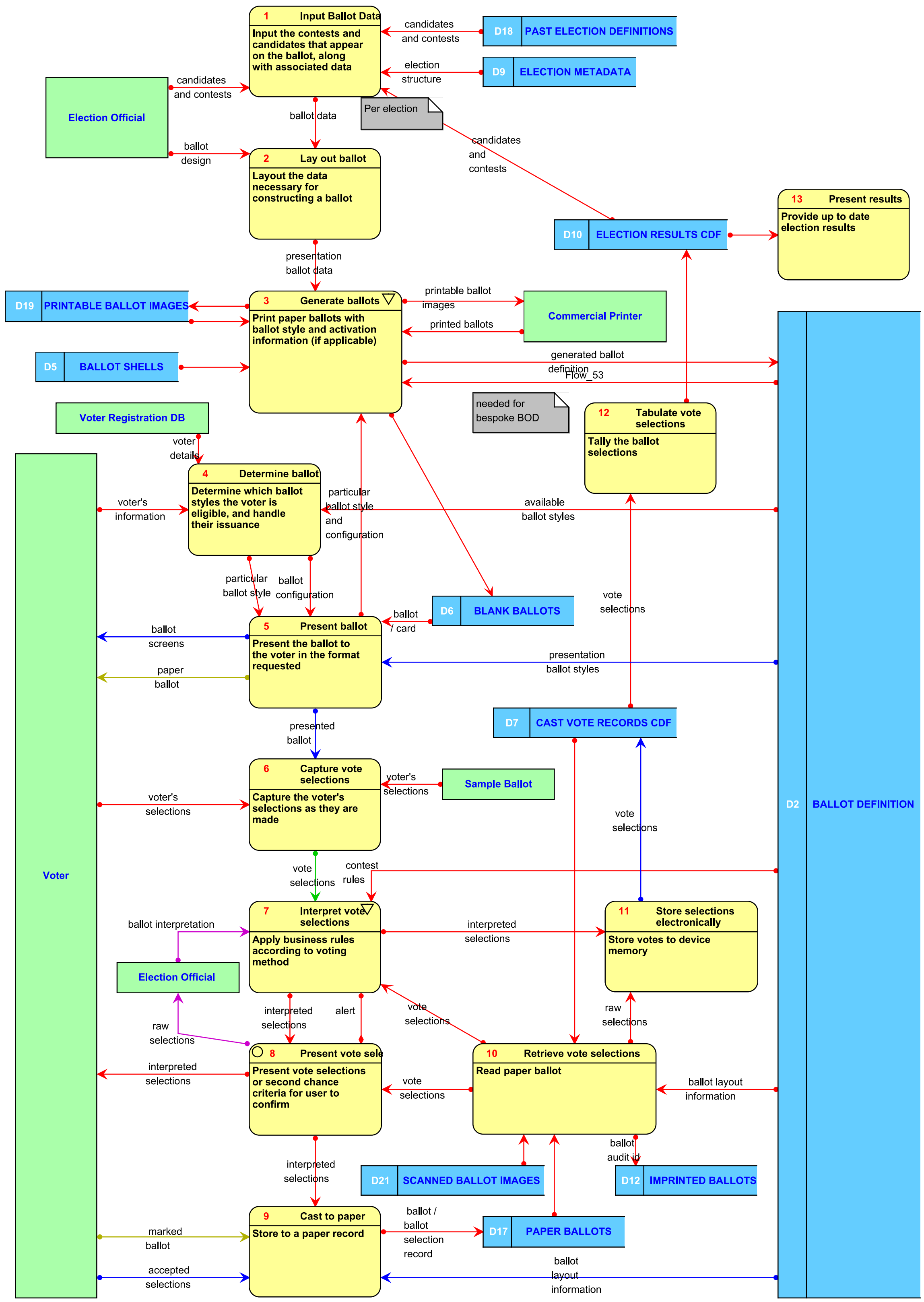


Figure 2 – Data flow diagram of functions in a voting system

A detailed treatment of each of the thirteen functions follows.

2.1. F1 Input Ballot Data

To seed the election management system with data, various data about an election must be entered into it. These include ballot data - things such as *contests, candidates, and associated geopolitical data*. This pre-election data can be transferred from other systems using the pre-election use-case of the Election Results Reporting (ERR) Common Data Format[4]. Customary practice is to clone data from a past election to speed up this task.

2.2. F2 Layout ballot

Each ballot style is laid out by putting candidates and contests (input in F1) in ballot order. Additionally, structural and design details, such as headings and other boilerplate, may be added.

2.3. F3 Generate ballots

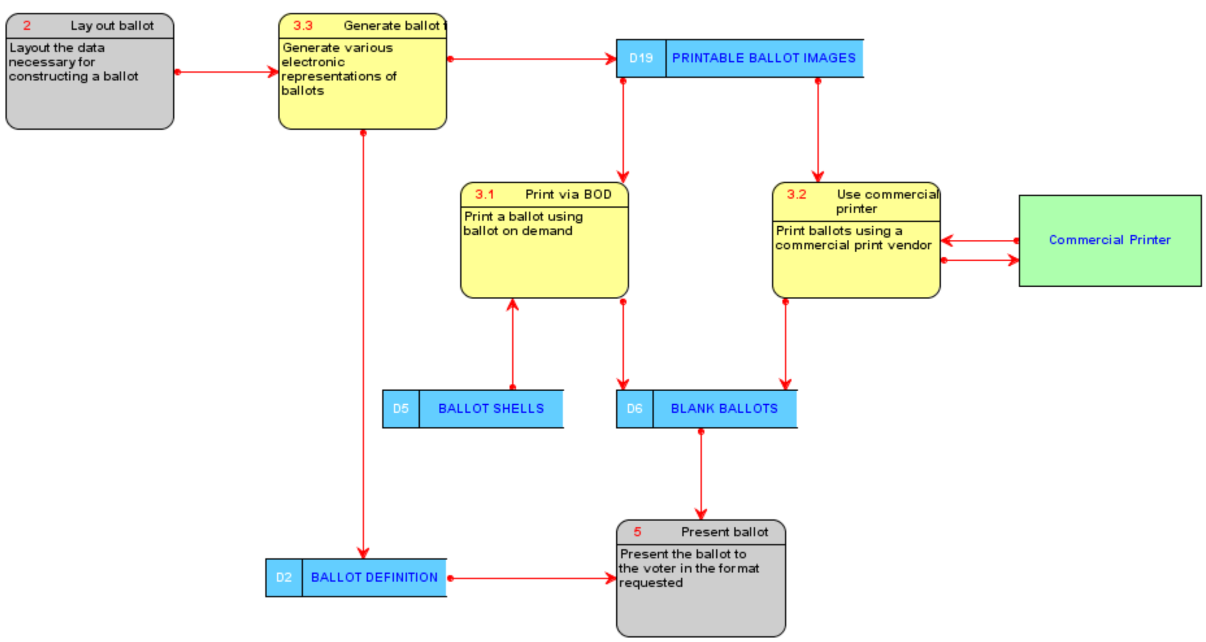


Figure 3 - Data flow diagram of generate ballots process

Ballots are generated electronically using the *presentation ballot data* (from F2), fixing the positions of contests, contest options, and contest option position appearances.

For ballots that will be displayed on a ballot marking device (BMD), this function's output is limited to the creation of a ballot definition file, which will feed the ballot marking and scanning devices, among others.

For full-face paper ballots, *ballot images* (e.g., a PDF) can be generated to share the print layout between systems and manufacturers. Such images can feed systems that cannot interpret a ballot definition directly, such as a third-party commercial printer. This function also handles the physical printing of the ballot, for voting systems capable of directly printing ballots (e.g., on-demand ballot printers).

2.4. F4 Determine ballot

Determining the ballot for a given voter involves *information about the voter*. This includes where that voter lives, their party affiliation, as well as accessibility requirements. Both address and party affiliation influence the *particular ballot style* that is provided, while the *voter's situation* may influence the configuration of the ballot, such as receiving one in larger print, or a particular natural language. The set of *available ballot styles* come from the ballot definition. Additional details may come from a voter registration database.

2.5. F5 Present ballot

Presentation of the ballot is highly dependent on what will be used to mark the ballot. In the case of hand marked paper ballots, the presentation of the ballot is simply the conveyance of the *paper ballot* from the election official to the voter. In the case of a ballot to be marked using a ballot marking device (BMD), a paper *blank ballot*, card or activator is provided. In the case of an on-demand ballot printing system type scenario or a BMD requiring a card preprinted with activation information, the particular ballot style required for the voter may be forwarded on to a system to *generate the ballot* (F3).

Even though paper is eventually involved in all cases, a BMD will present screens to the voter. These screens serve as the surface in which the voter indicates their desired selections.

2.6. F6 Capture vote selections

In a ballot marking device (BMD) scenario, the capture of *vote selections* refers to the interaction of the voter and the BMD. It relies on *presented ballot* data retrieved from the previous function (F5) to overlay the *voter's selections* atop the contest options, and thus is functionally dependent on *present ballot*.

Voters may use the touchscreen and/or physical buttons of the BMD to make their selections. Voters' selections may come from an Interactive Sample Ballot, containing selections made in advance, often stored on a barcode.

In the case of HMPBs, "Capture vote selections" simply refers to the voter marking the ballot.

2.7. F7 Interpret vote selections

Vote selections are interpreted according to rules associated with the ballot and the voting method of each contest. For example, a selection in a straight party contest may cause indirect selection(s) in other contests. Additionally, one or more contests may be determined to be undervoted or overvoted according to particular voting method rules. In the case of a ballot marking device *in an active voting session*, selections may come directly from other functions (e.g., F6). In other cases, vote selections will come from a device that has retrieved them (F10), e.g., from a paper ballot via a scanner, or cast vote record.

Interpretation of a different kind can take place, in the adjudication of voter intent. This involves an election official reviewing the marks on the ballot and making determinations on whether those marks constitute selections for one or more available contest options. In this case, the election official is presented with the vote selections (F8) before final interpretation takes place.

Depending on the configuration of the interpreting component and the outcome of the interpretation, this information may be forwarded to *present vote selections* (F8) for confirmation, or directly stored to device memory (F11).

2.8. F8 Present vote selections

Vote selections are presented to voters to confirm them prior to casting, or presented to election officials after casting to confirm voter intent. Presentation of vote selections may be a step in a voting session on a BMD. Vote selections may also be presented to the voter after marking by another ballot marking device or voter-facing scanner. This is often a consequence of an alert condition caused by how the ballot was marked (e.g., presenting a contest and contest option selections (if any) that led to an overvote or undervote).

In a BMD scenario, presenting vote selections (F8) is the last step before the *accepted selections* on the ballot are *cast to paper*. For scanners, presentation of vote selections may be the last step before the *accepted selections* on the ballot is stored to electronic representation (i.e., an electronic cast vote record).

2.9. F9 Store to paper

Selections made on a ballot marking device must be stored to paper to meet software-independence requirements. For hand marked paper ballots (HMPB), this function is implicitly performed.

2.10. F10 Retrieve vote selections

Retrieval of vote selections occurs at or after casting, in order to perform tabulation and other prerequisite activities. A paper ballot is interpreted on first retrieval (F7).
For a ballot that has been scanned previously, vote selections can be retrieved from a cast vote record. Retrieved vote selections can be passed down to another function to store them electronically, without first interpreting them (F11), or sent to be interpreted (F7) or presented (F8).

2.11. F11 Store selections electronically

Selections that have been read (but not interpreted, F10) or interpreted (F7) can be stored to electronic memory (in forms such as the Cast Vote Record CDF[2]). Retrieved or interpreted selections come from scanners.

2.12. F12 Tabulate vote selections

The tabulation of *vote selections* is supported by the Cast Vote Records Common Data Format.

2.13. F13 Present results

The presentation of results is the creation of reports for human consumption, either on paper or through electronic interfaces.

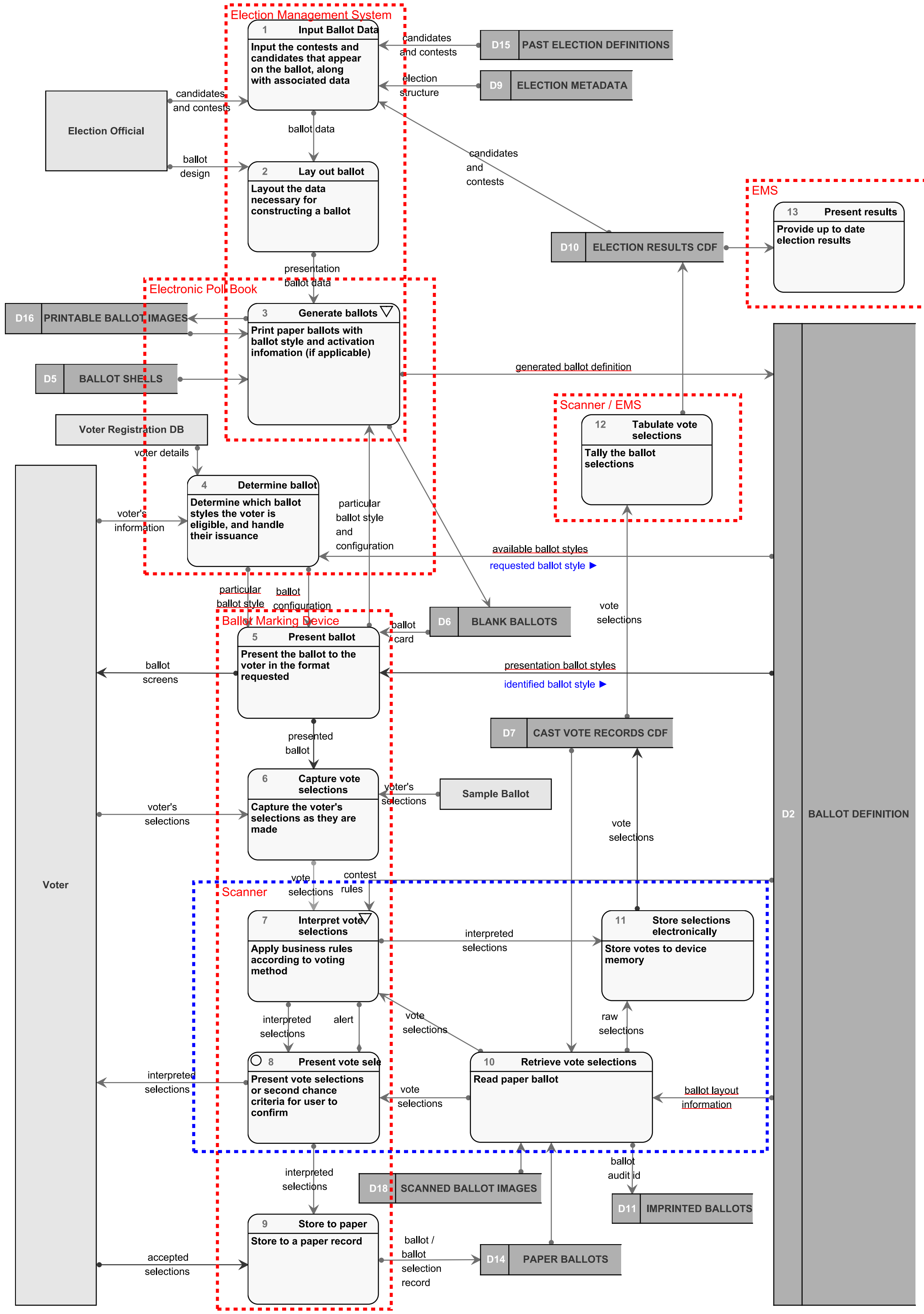
3. Interoperability Scenarios

This section describes a set of common interoperability scenarios. Each scenario includes the data flow diagram from Section 2 with the components performing each function laid on top (as boxes with dotted lines). The data flows between component boundaries that are identified as significant for interoperability are underlined.

Potentially significant data flows include any that cross the boundary of a component. However, the determination of what data flows are *significant* for interoperability are made based on the need for machine readable data between components. Data flowing informally (e.g., via a human via a user interface) are out of scope.

Significant data flows are grouped together based on the data they convey. An analysis for each unique data flow is given in Section 4.

3.1. S1 Ballot Marking Device with Voter-facing Ballot-Aware Scanner



This scenario uses the following election technology components:

- Election Management System
- Electronic Poll Book
- Ballot Marking Device
- Voter-facing Ballot-Aware Scanner

The election management system performs functions F1-F3. F3 creates the *ballot definition*. An electronic poll book uses the created *ballot definition* to receive a list of *available ballot styles*. Data from a voter registration database contains *details about individual voters*, such as precinct/district assignment and eligibility information. These data sets are used by the EPB to determine which ballot (F4) the voter should receive. A ballot marking device performs F5-F9. The BMD relies on receiving the particular ballot style (as an identifier, likely from the EPB) that the voter has chosen, along with configuration data that sets up the voting session. Presentation ballot style data required to render the ballot are provided by the ballot definition, separately, likely during device loading (via the Election Management System).

It is possible that the voter brings an interactive sample ballot containing their selections. This can be used to capture vote selections (F6).

At the ballot *cast to paper* step, the ballot selection record or machine-marked full-face ballot (in either case, paper) is generated. This is fed into the scanner (performing F7, F8, F10, and F11). The scanner must also have access to ballot layout information, which it needs in order to interpret the encoding of the selections. Depending on interpretation of the ballot, the cast vote record is directly generated, or presented for the voter for confirmation first.

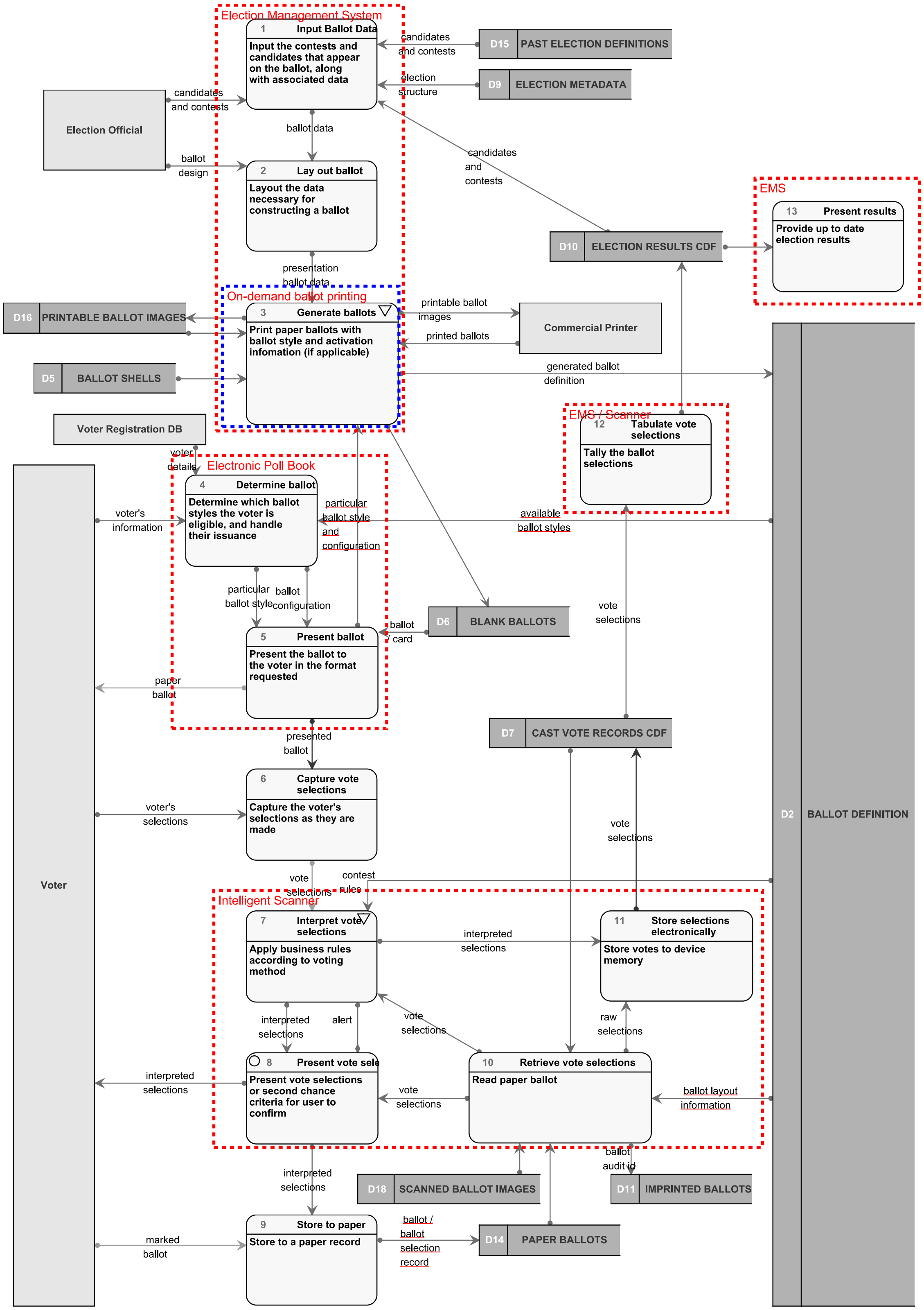
(F12 and F13 do not vary between scenarios and do not use ballot definition information)

Table 1 - Significant Data Flows

Between Component Boundaries of	**Flows**
Election Management System and Electronic Poll Book	• Available ballot styles
Electronic Poll Book and Ballot Marking Device	• Particular ballot style[a] • Ballot configuration[a]
Election Management System and Ballot Marking Device (via CDF)	• Presentation ballot styles
Ballot Marking Device and Scanner (via paper)	• Ballot / Ballot selection record
Election Management System and Scanner (via CDF)	• Ballot layout information
Voter Registration Database and Electronic Poll Book	• Voter details

[a]Assumes EPB performs encoding

3.2. Hand Marked Paper Ballots with On-Demand Ballot Printing (ODBP) and Ballot-Aware Scanner



This scenario uses the following election technology components:

- Election Management System
- Electronic Poll Book
- On-demand ballot printing system
- Ballot Marking Device
- Ballot-aware Scanner

The election management system performs F1-F3. F3 creates the *ballot definition*. An electronic poll book uses the created ballot definition to receive a list of *available ballot styles*. Data from a voter registration database contains *details about individual voters*, such as precinct/district assignment and eligibility information. These data sets are used together to determine the ballot (F4) the voter should receive. An on-demand ballot printing system device (separate from the electronic poll book) produces the paper ballot (F3), using ballot images previously generated by the EMS (also performing F3).

The paper ballot is used to capture vote selections (F6). (F9 is not performed because the ballot is already paper). A ballot-aware scanner is used to retrieve the vote selections (F10). This requires access to the paper ballot as well as the ballot layout information. The ballot layout information is used to correlate the ballot style and interpret the vote selections (F7). The vote selections are conditionally presented to the voter (F8). An accepted ballot is stored electronically as a cast vote record (F11).

(F12 and F13 do not vary between scenarios and do not use ballot definition)

Table 2 - Significant Data Flows

Between Component Boundaries of	**Flows**
Election Management System and Electronic Poll Book	• Available ballot styles
Electronic Poll Book and On-demand Ballot Printer	• Particular ballot style • Ballot configuration
Election Management System and On-demand Ballot Printer	• Presentation ballot styles
On-demand Ballot Printer and Scanner (via paper)	• Ballot / Ballot selection record
Election Management System and Scanner (via CDF)	• Ballot layout information
Voter Registration Database and Electronic Poll Book	• Voter details

4. Identified Cross-Component Flows

This section describes in detail the data required for each data flow identified as significant for interoperability.

The scenarios enumerated in the previous section have demonstrated the following flows as significant for interoperability. For purposes for this discussion, these flows have been generalized to:

Significant Data Flows	**Generalized as**
Available ballot styles	Ballot creation/activation information
Particular ballot style (identifier)	
Ballot configuration	
Presentation Ballot Styles	Ballot Layout
Ballot / Ballot selection record / Ballot Layout	Voter's contest selections

Unlike earlier sections that described data flows between functions, we will describe each significant flow in terms of voting system components.

4.1. Ballot creation/activation information

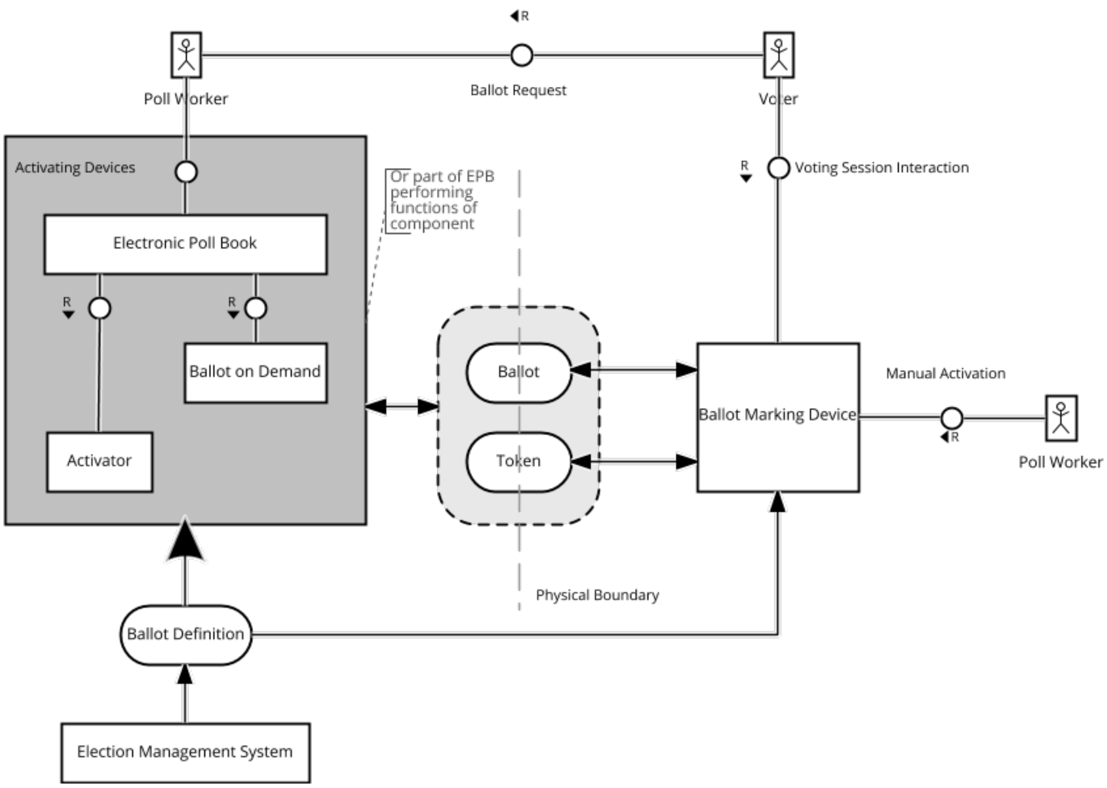


Figure 4 - Block diagram of components involved in ballot activation

After it is determined which ballot style a voter should receive (F4) and its associated configuration (likely via electronic poll book), a ballot of that style must be presented to the voter.

In the case of a hand-marked paper ballot configuration with preprinted ballots, this is as simple as poll workers locating the correct ballot style, possibly recording which unit of ballot stock was used (e.g., from a serial number on a stub), and handing it to the voter. In an on-demand ballot printing (ODBP) scenario, the identified ballot and associated configuration can be forwarded from the EPB directly to a printer, or ODBP software (F3).

In the case of a ballot marking device (BMD) scenario however, the activation device must be able to determine which ballot style to encode, and in addition, any contextual details about the voter's situation so that the voter receives a ballot accessible and usable to them. Lastly, activation of a voting session must be secure, so that only "activated" ballots can be marked on a BMD.

The constraints put on activation include a restriction on computer networking between the voting system and other components. Particularly, electronic poll books are often connected to wide area networks (WANs) and cannot connect to voting system components.

Possible BMD activation scenarios are described below:

4.1.1. S1 Blank full-face ballot stock

A full-face paper ballot is provided to the voter. The voter inserts the ballot into the BMD. The BMD performs the ballot style identification optically, by reading preprinted impressions on the ballot. The ballot is rendered via screens on the device.

4.1.2. S2 Activation card

An activation card (which may or may not become the eventual ballot selection record) is activated by printing specific information (likely as a barcode) onto it. The card is then provided to the voter. The voter inserts or scans the card into the ballot marking device. The BMD performs the ballot style identification optically by reading markings on the card. The ballot is rendered via screens on the device.

4.1.3. S3 Activation token

An activation token (e.g., RFID) is programmed with specific information for the voter's voting session. The token is then provided to the voter. The voter inserts or otherwise interfaces the token with the ballot marking device. The BMD performs the ballot style identification electronically. The ballot is rendered via screens on the machine.

4.1.4. S4 Manual poll worker activation

In this scenario, the ballot activation is handled by the poll worker by configuring the ballot marking device with the correct parameters for the voter's chosen ballot style and situation.

4.1.5. Analysis

It should be possible to create a common data format that supports the activation of ballot marking devices and the creation of ballots (e.g., via ODBP).

Because of air gap requirements between the voting system and other devices (e.g., the electronic poll book), the use of an electronic activation media (e.g., token) or paper card may be required.

To interoperate between components performing ballot activation and ballot marking, both the data format and the medium used to convey that data are significant. When electronic activation media is used, the nature of the electronic interface is significant. When paper is used, the encoding of the activation data on paper (e.g., a barcode) is significant.

Some paper activation cards may be used as the storage medium of ballot selections (thus also serving as the ballot selection record). This effectively avoids the issue of activation cards being reused, as it should not be possible to turn a single card into multiple ballot selection records. However, for ballot marking devices that contain their own stock to produce ballot selection records, the risk of the reuse of ballot activation cards to create more than one voting session must be considered.

The following is a non-exhaustive list of information that may need to be exchanged to support *ballot creation and activation*.

- required ballot style (identifier)
- party affiliation (closed primaries)
- accessibility information (screen size / text-to-speech / assistive technology (AT) info)
- chosen natural language

The exchange must be done in a manner that ensures data integrity and authenticity. Additionally, the exchange must be certain not to directly or indirectly contain data that could be used to identify a particular voter.

4.2. Voter's contest selections

Contest selections captured on a paper ballot or paper ballot selection record must be scanned before they can be tabulated. In order for ballot producing devices (including EMSs and BMDs) and scanners from different vendors to interoperate, a mutual understanding of how contest selections are represented on a software-independent medium (e.g., paper) is required.

Scanners require two related sets of information about ballot styles. The first is general information about how data is encoded on a ballot. This could be instructions on how to decode the location and format of a ballot identification code (in an OMR ballot using identification codes) or an equivalent textual representation (for an OCR ballot). Such data would be used by the scanner to identify the ballot style being scanned and locate the second piece of information, its associated contest response layout. The contest response layout describes how contest selections are captured on the ballot.

4.2.1. Sources of data for ballot styles

Ballot data can come from a variety of sources. Most often it is stored within a candidate qualification subsystem of a voter registration database (VRDB). It is then fed downstream to other systems, such as a voting system's election management system (EMS). The EMS then creates the ballot definition that can be used by other components of the voting system. This is the standard approach for voting systems whose components are certified together.

However, in order to support componentization within voting systems, devices from different manufacturers, such as those that generate ballots (F3) and those that retrieve vote selections (F10) must be able interoperate using a ballot definition in a common data format. Particularly, a ballot definition must contain a contest response layout specifying how contest selections are captured and where they can be located. Figure 5 describes a voting system where the system generating ballots ("Full Face Ballot System") and the scanner consuming those generated ballots are produced by different vendors.

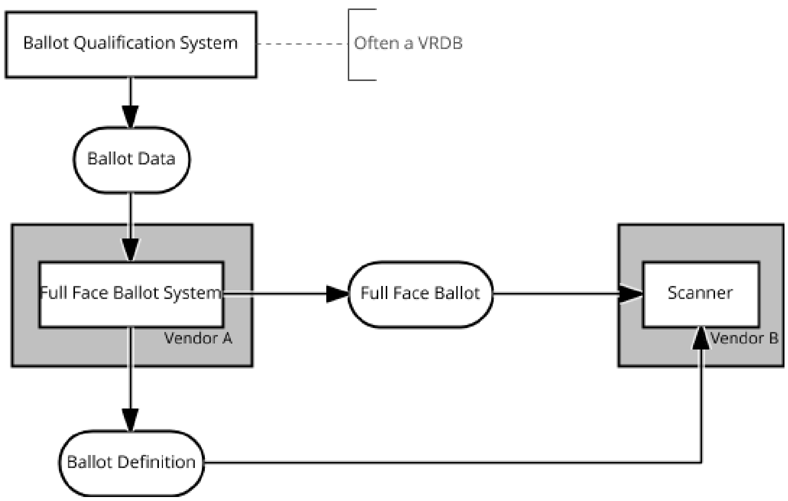


Figure 5 - Depiction of a voting system scanner and EMS from separate vendors

Systems that support scanning multiple, differently sized paper vote records are becoming more popular as voters receive full-face ballots via postal mail, and in-person voters use ballot marking devices that print onto ballot selection records. Some election jurisdictions offer remote ballot marking, which provides a similar interface to that of a BMD in a home or other remote setting and prints vote records onto standard (i.e. ANSI/ISO) paper sizes.
For a single ballot definition to serve multiple ballot representations, e.g., full-face paper ballot and card producing BMDs, the CDF will need to contain representations for each, and a method to distinguish them when scanned, so that correct contest selections may be located. In the following example, a single ballot definition is shared between a system capable of generating full-face ballots and an EMS that is included with a BMD system. In this scenario, it is expected that the ballot definition would have some means to separate the details related to the ballot selection record format used (as output of BMDs) and that of a full-face paper ballot.

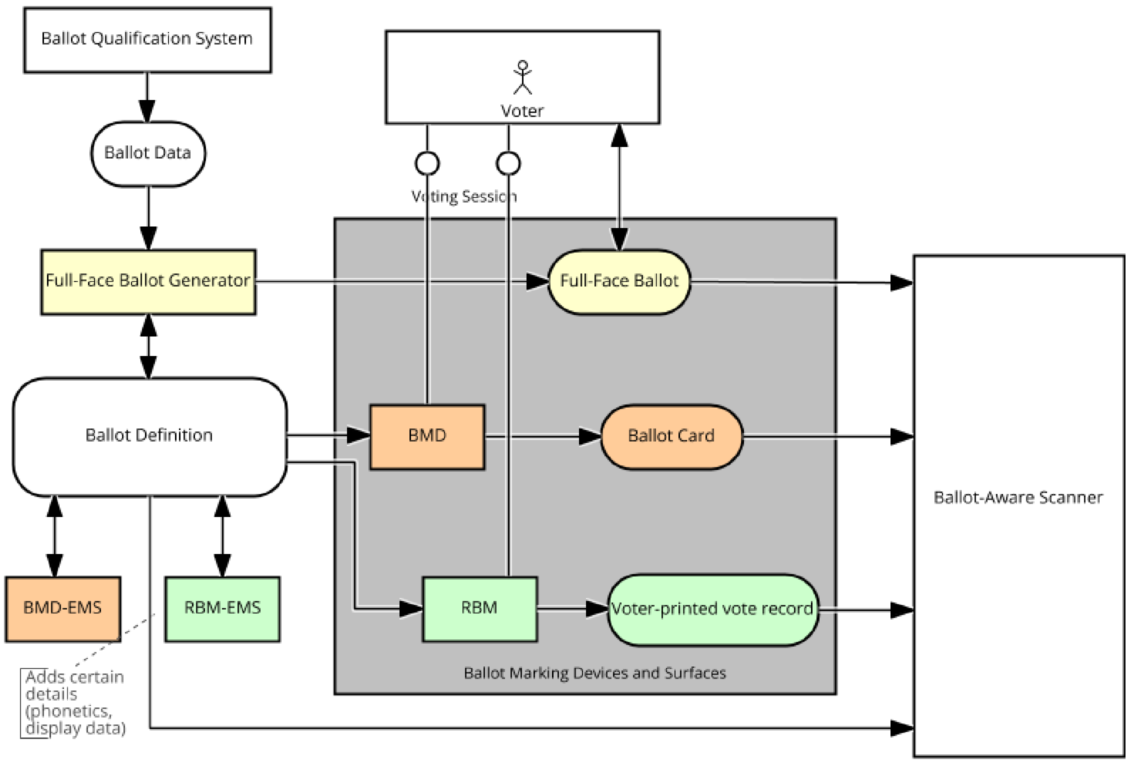


Figure 6 - Ballot definition shared between multiple components of different vendors

Alternatively, the Full-Face Ballot Generator and Ballot Marking Device could use separate ballot definitions and scanners (Figure 7). However, in order for results derived from both full-face paper ballots and ballot selection records to be tabulated and aggregated together, contests and associated options and selections will need to possess shared identifiers, which would come from the upstream system providing ballot data.

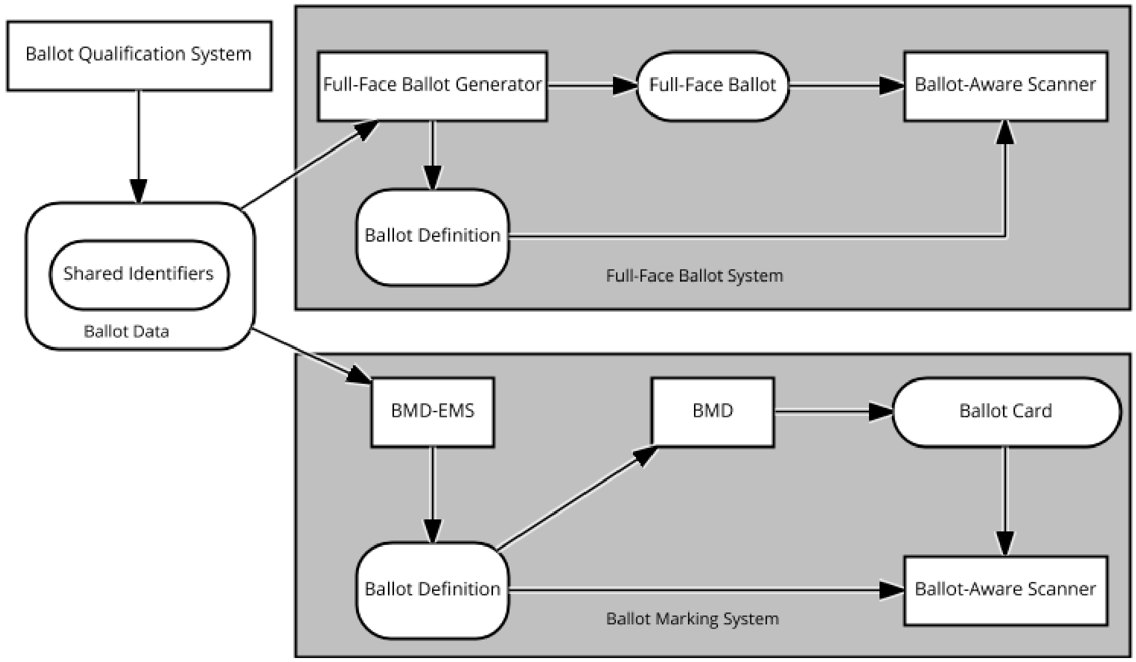


Figure 7 - Separate scanners for each ballot medium

4.2.2. Significance of marks

The parts of the ballot that are relevant for interoperability vary based on how the scanner detects the contest selections on the ballot. For example, a scanner that captures contest selections via optical character recognition (OCR) may consider the textual content used to represent contest options and the contest option position indicator appearance (e.g., oval, broken arrow, etc.) used.

A scanner that captures contest selections using optical mark recognition (OMR) may need to know the relative position of contest option position indicators only. The following section describes different scanning methods in detail and how each influences its interoperability needs.

4.2.3. Scanning Technologies

Optical mark recognition is the dominant method of reading ballots. Traditional OMR scanners contain optical sensors (read heads) placed in fixed positions associated with areas where a mark could appear.[1] Newer devices use imaging sensors (e.g., Charge-coupled device, contact image sensor), similar to those of a digital camera, capable of producing electronic images of the printed ballot, a "ballot image". This enables scanning

[1] https://www.pilotltd.com/en/blog/how_omr_works

using techniques such as optical character recognition in addition to optical mark recognition. Further, the creation of ballot images enables additional use-cases, such as electronic adjudication of ballots (F7), including adjudication of write-ins, and may aid in auditing.

Both OMR and OCR are capable of reading barcodes. Barcodes have varied purposes on ballots but are used in particular as output of a ballot marking device voting session, i.e., as the machine readable portion of a ballot selection record.

4.2.3.1. Optical Mark Recognition (OMR)

An OMR ballot contains contest option positions in predetermined areas. During scanning, a “master barcode”, in a known position, is used to locate the “codebook” within the scanner’s storage that can be used to decode the meaning of other marks on the ballot. OMR can be performed with traditional mark sensing optical sensors (e.g., LEDs) or imaging sensors (e.g., CIS, CCD, CMOS).

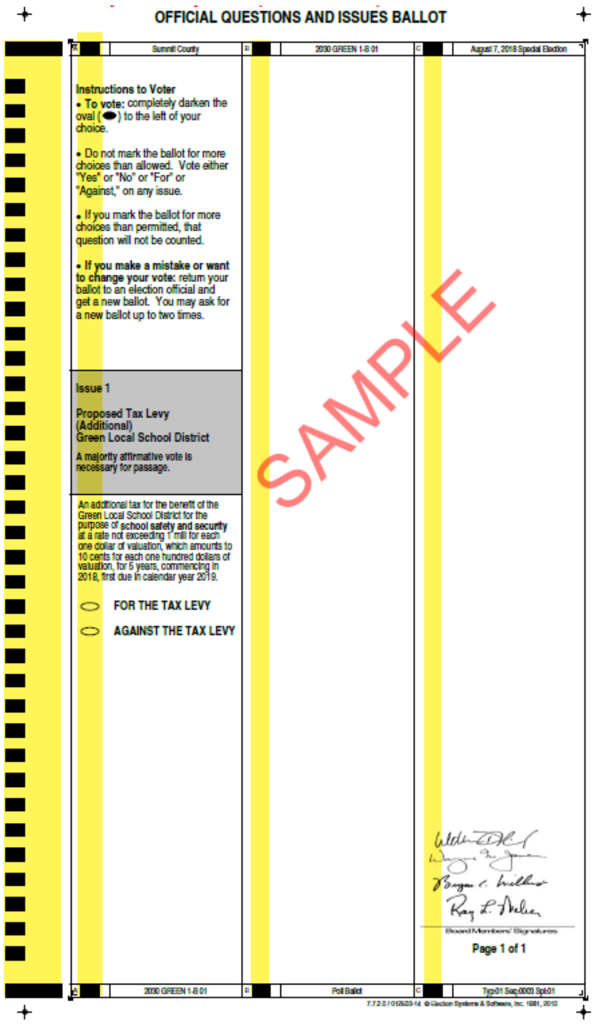


Figure 8 - Ballot with areas significant for OMR highlighted

Paper size, margins, and aspect ratio

Traditional OMR systems have strict requirements for paper size, margins and spacing. This is because the placements of optical sensors are physically fixed. Paper sized differently than specified could lead to marks appearing at locations on paper not expected by the scanner. Modern systems are somewhat more lenient, but these factors still play into readability.

The positional grid pattern on the ballot

Modern OMR ballots generally contain two sets of tracks composed of marks (usually rectangles) that form one or more rows (indicated by timing marks across a track) and columns (channels). These two tracks create a virtual two-dimensional grid showing all possible locations where a mark could appear on that ballot. An interoperable scanner would need to be able to recognize this grid and marks placed on it.

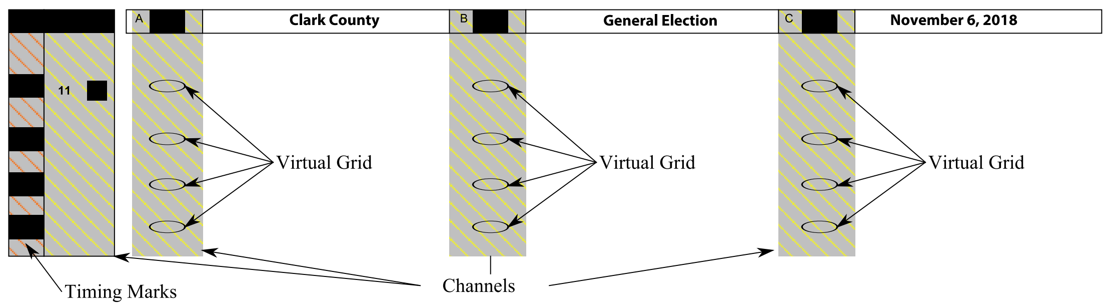


Figure 9 - Virtual grid of a ballot scanned using fixed-position sensors

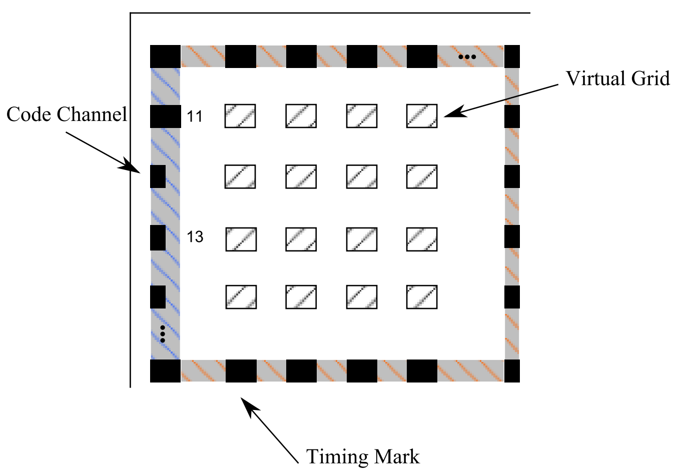


Figure 10 - Virtual grid of a ballot using digital imaging

(Not all ballots that use OMR require the use of timing marks.)

Contest Option Positions and Appearance

The *contest option position*, informally called a target area, is an enclosed portion of the ballot where a selection can be made. Contest option positions constitute a subset of positions on the grid (see above). Associated with this position is an *indicator*. Indicators are visual representations, often ovals, boxes, broken arrows, or other simple geometry. These provide a visual cue to the voter as to where a mark for a particular contest option is expected.

Codebook

Because the ballot and selections are encoded in a particular way, shared knowledge of this encoding between systems printing ballots (F3), ballot marking devices and scanners is vital. A codebook contains a mapping of contest option positions to their underlying contest and candidate data. For example, the position represented by the cartesian coordinates (3,4) on the ballot could correspond to the contest option George Washington, a candidate for contest of President of the United States.

Shared Standards

For the BMD and Scanner to communicate effectively, both the BMD and the scanner must interpret the ballot in the same way. For example, if the minimum mark density of a ballot marking device is lower than that of a scanner, marks may not be properly interpreted as contest option selection.

Analysis

Areas of interest to CDF development would include specific details for each area listed above. It should be possible to describe a grid in terms of number of columns and rows per page. The physical artifacts of the grid should also be expressible as data (offset, thickness, width, etc.), and the bounding box (extent) of the grid. Likewise, it should be possible to express the *contest option position indicators* in a CDF, using standard geometric primitives (e.g., lines, circles, polygons, etc.).
Because almost no contemporary equipment uses traditional fixed position scanning, this use-case should not be considered a constraint for common data format development. Modern image processing techniques should be able to handle paper size and image scaling gracefully.

- Data Points of Interest
    - Mark type (timing control, direct under type, mark to mark, and FACOM)
    - Mark width
    - Number of columns
    - Spacing of timing marks

4.2.3.2. Barcoded Ballot Selection Record

Ballot selection (or "summary") records often use barcodes to represent selections on a ballot. These "ballots" usually only contain the contests and contest options that have been selected by the voter during the voting session. One or more barcodes are used to encode the contest selections and other human readable information. Two-dimensional barcodes (e.g., a QR Code) are able to store more information than a standard barcode and are commonly used.

```
BALLOT SELECTION RECORD (SPECIMEN)
TEST ELECTION
11/2/2021
PRECINCT 1

FAVORITE ICE CREAM ____________________________________
                                                VANILLA
FAVORITE ANIMAL _______________________________________
                                                  ZEBRA
FAVORITE VOCAL ARTIST _________________________________
                                             THE SONICS
                                       GIL-SCOTT HERON
INCOME TAX HOLIDAY ____________________________________
                                                    YES
LIBRARY RENEWAL LEVY __________________________________
                                                    YES
```

Figure 11 – Example “Selections-only” Ballot Card

Most barcodes support laser based as well as modern OMR and OCR scanning.

Many factors affect the reading of a barcoded ballot. Those include:
Symbology used (e.g., Code 128, QR Code, PDF417, etc.)

Symbology refers to the size, shape and arrangement of symbols that can constitute a barcode. Many different barcode symbologies have been standardized.

Quiet zone

Barcodes and scanners have different tolerances for reading barcodes surrounded by other content.

Error detection and correction

Barcodes may have built-in error correction or detection which can be used to decode partially damaged or misprinted barcodes.

Barcode content encoding

The content of the barcode may be encoded in different ways to best utilize their limited capacity. Barcode types vary in the content they can encode. Some can encode arbitrary binary content, while others may be limited to a subset of characters.

Barcode content

Currently, the barcode content may vary from manufacturer to manufacturer.

Analysis

While many barcodes use symbologies that are largely standardized, the data they contain is not. Although many current ballot designs use non-standard and proprietary barcode designs that are not publicly disclosed, VVSG 2.0 requires use of standard barcode designs. Different symbologies can be used to encode the same data. For a BMD and scanner to interoperate, the scanner must be able to decode not just the symbology into its machine-readable content (i.e., a text string or binary), but also its meaning. This often involves encoding the selections using codes and referring to a codebook to decode them. Thus, to fully understand the meaning of the barcode, it must be decoded twice, once from its visual symbology, and again from its machine-readable encoding.

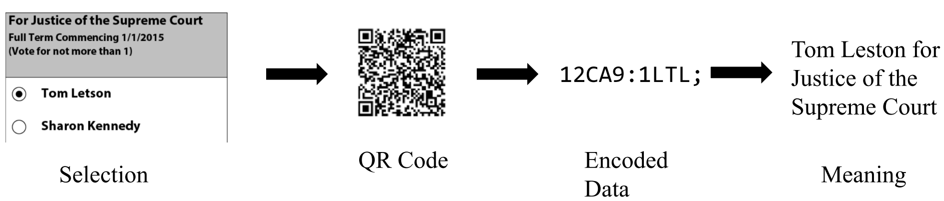


Figure 12 - Semiotic transformations of selections

4.2.3.3. Optical Character Recognition

Last for consideration is Optical Character Recognition (OCR). OCR relies on a digital image having been captured by the scanner, and then performing analysis on the image to determine its content. OCR distinguishes itself from OMR in that in addition to reading marks designed for machine readability (e.g., barcodes), it is also able to read typography (and potentially handwriting). All ballots that can be read by an OMR system can theoretically be read by an OCR system.

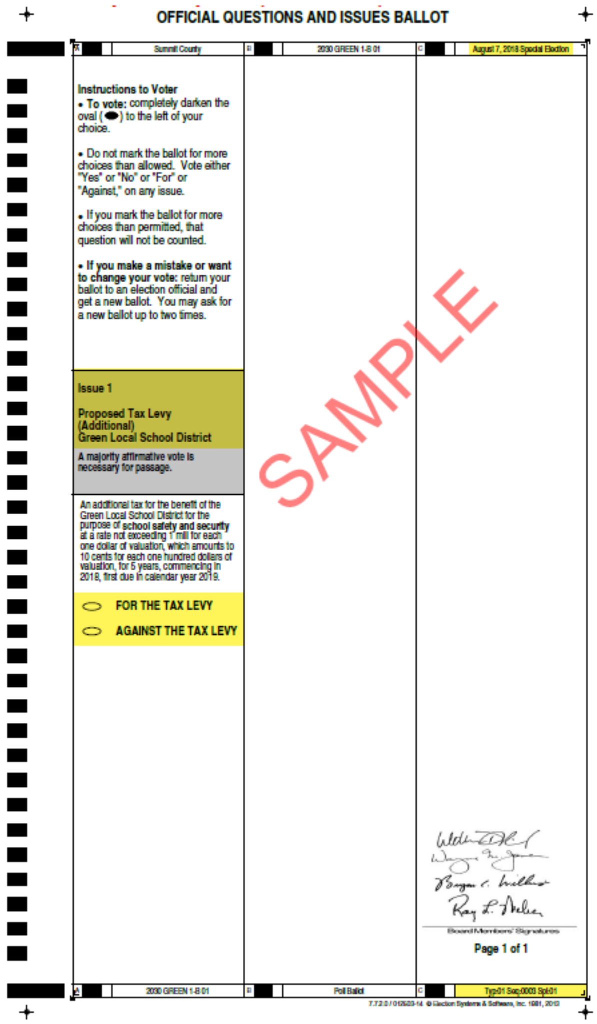


Figure 13 - Ballot with areas significant for OCR highlighted

Many factors affect reading an OCR ballot. Those include:

- Font type, size, color
- Relative position of ballot content (i.e., position of contest and candidates)
- Ballot layout (including contest breaks across content areas or pages)

4.2.3.4. Analysis

OCR has been slow be used in ballot scanning applications. Unlike OMR, complex pattern recognition is required, and involves additional processing power. Because OCR uses human readable text, it will be heavily dependent on the layout of the ballot (see Section 4.3).

4.3. Ballot Layout

Ballot layout refers to the aspects of the ballot the voter is intended to perceive and understand. This includes instructions on how the ballot may be voted, headers, contests and associated options. Ballot layout requirements are often set in statue and may constrain the layout's fonts, point sizes, contest and page breaking, among other items.

4.3.1. Analysis

The first common data format in this area was IEEE 1622-2011[5] which was a profile of the Election Markup Language (EML) [6] for blank ballot delivery systems. An effort started in 2017 on ballot definition led to the addition of headers to the Version 2.0 of the Election Results Reporting (NIST SP 1500-100) [4]. That group did not consider presentation details required to produce full-face paper ballots.

Ballot layout details specific to a particular presentation method, such as electronic or paper must be taken into account. It is conceivable that an EMSs might produce ballot definitions lacking enough detail to render a ballot using election technology components. For example, an EMS designed primarily for paper ballots may not store information about spoken phonetics used by the assistive technology (AT) for a BMD.

The Pre-Election use-case of the Election Results Reporting CDF contains several of the structural data required to support ballot layout and could be leveraged as a starting point.

5. Conclusions

We have found a significant need for ballot style interoperability. Ballot styles open up the capability for component level certification of devices including ballot marking devices and scanners.

5.1. Ballot styles interoperability is highly complex

A ballot styles CDF might be able to support interoperability between ballot marking devices and scanners, but it is not just a matter of constructing a machine-readable format. In fact, in order to interoperate between the BMD and scanner, the ballot itself must be interoperable.

To have an interoperable ballot the form of the ballot must be describable by the yet to be defined Ballot Styles CDF. To achieve this level of interoperability, we must consider not just data, but function. How a machine responds to certain data is as important as the form of the data itself. This opens up new questions to explore in future work.

5.2. Recommendations

5.2.1. Consider standardizing use of barcodes within voting systems

Our analysis found multiple places where barcodes are used in the voting process. Barcodes are used to communicate between systems and are thus of particular interest for interoperability. Particularly, they are used between EPBs and BMDs to produce an activated ballot, and between BMDs and scanners to store voted ballot data including contest option selections. Research should be done on existing information encoding techniques currently used on barcoded ballots and activators and consider if a standard for information encoding is warranted.

5.2.2. Develop a Ballot Styles Common Data Format

This paper shows a definite need for a ballot styles CDF. Use-cases for such a standard are thoroughly addressed elsewhere but to restate:

- Support the description of ballot layout
- Support ballot creation and activation
- Support conveying information necessary for a scanner to read a voted paper ballot produced by a BMD or full-face ballot producing component.

5.2.3. Ensure Ballot Styles CDF supports usability and accessibility best practices (Human Factor Considerations)

Ballot styles deal extensively with presentation, and in several cases paper. It may be useful to determine if usability and accessibility best practices could be leveraged to make the layout of ballots more predictable, and thus ease the interoperability of reading physical ballots.

Human factors considerations include:

- Support for best practices in ballot layout
- Protect the property of "marked as intended"
- Provide usability and accessibility guidance around standards for ballot activation

5.2.4. Security Considerations

Security considerations include:

- The use of certain encodings on a ballot (e.g., barcodes)
- Ballot activation
- Protection of the property "counted as cast"
- Maintaining data integrity during data exchange
- Determining the integrity of input ballots and encoded media.

5.2.5. Ensure the testability of Ballot Styles CDF and Interoperable Ballots

Because ballot style interoperability requires more than just data interoperability, new test methods may need to be developed to verify conformance. Involvement of the testing community, including Voting System Test Labs (VSTLs) and the Election Assistance Commission (EAC) will be crucial to this effort's success.

Acknowledgments

We would like to acknowledge the members of the NIST Common Data Format Research Group who lent their time and expertise to this document.

References

[1] Gane, Chris P. and Sarson, Trish (1979) Structured Systems Analysis: Tools and Techniques (Prentice Hall Professional Technical Reference), 1st Ed.
[2] Voluntary Voting Systems Guidelines, Version 2.0, February 10, 2021, Available at https://www.eac.gov/sites/default/files/TestingCertification/Voluntary_Voting_System_Guidelines_Version_2_0.pdf
[3] Wack, John (2019) Cast Vote Records Common Data Format Specification. (U.S. Department of Commerce, Washington, D.C.), NIST Special Publication (SP) 1500-103,Version 1.0.
[4] Wack, John (2019) Election Results Common Data Format Specification. (U.S. Department of Commerce, Washington, D.C.), NIST Special Publication (SP) 1500-100,Version 2.0.
[5] IEEE (2012)"IEEE Standard for Electronic Distribution of Blank Ballots for Voting Systems," in IEEE Std 1622-2011, doi: 10.1109/IEEESTD.2012.6130556.
[6] OASIS, Election Markup Language (EML) Specification Version 7.0, Committee Specification 01, October 27, 2011. Available at: https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=election [accessed 2/1/2019].

6. Acronyms

Selected acronyms and abbreviations used in this document are defined below.

Acronym	**Meaning**
ANSI	American National Standards Institute
AT	Assistive Technology
BMD	Ballot Marking Device
CCD	Charge-Coupled Device
CDF	Common Data Format
CIS	Contact Imaging Sensor
CMOS	Complementary Metal-oxide Semiconductor
CVR	Cast Vote Record
EEL	Election Event Logging
EMS	Election Management System
EPB	Electronic Pollbook
ERR	Election Results Reporting
HMPB	Hand Marked Paper Ballot
ISO	International Organization for Standardization
LED	Light-emitting diode
OCR	Optical Character Recognition
ODBP	On-demand Ballot Printing
OMR	Optical Mark Recognition
PDF	Portable Document Format
RBM	Remote Ballot Marking
RFID	Radio-frequency Identification
VRDB	Voter Registration Database
VRI	Voter Records Interchange
VSTL	Voting System Testing Laboratory
VVSG	Voluntary Voting System Guidelines
WAN	Wide Area Network