FVAP

DATA MIGRATION TOOL

A CSG OVERSEAS VOTING INITIATIVE REPORT

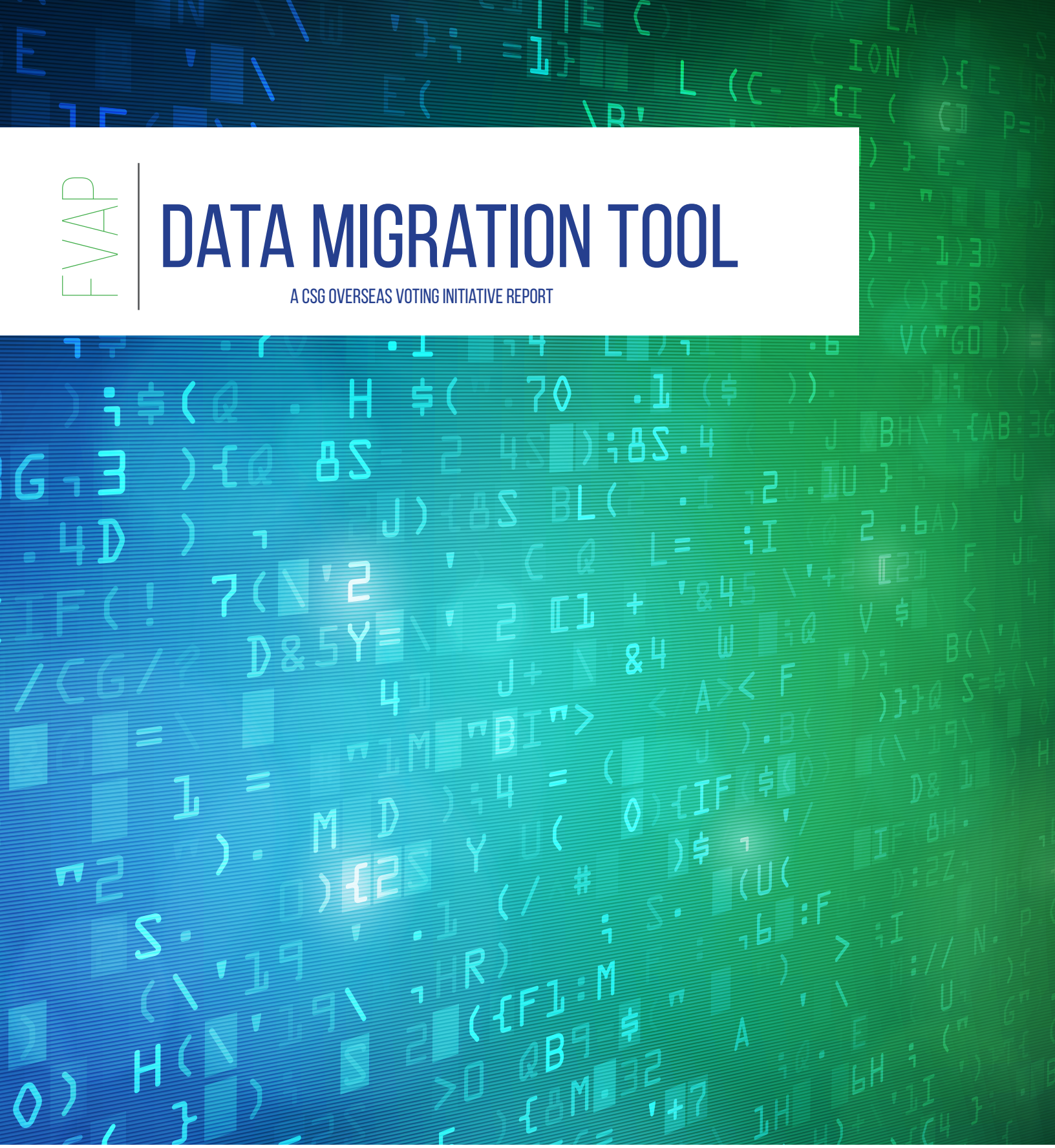


TABLE OF CONTENTS

INTRODUCTION

As part of its cooperative agreement with the Federal Voting Assistance Program of the U.S. Department of Defense, The Council of State Governments continues to research efforts in the area of data standardization. This effort is intended to help the Federal Voting Assistance Program, or FVAP, better understand its key policy and program areas and support data interchanges to assist state and local election officials deliver blank ballots to military and overseas voters in a flexible, efficient way. FVAP works to ensure military service members, their eligible family members and overseas citizens are aware of their right to vote and have the tools and resources to successfully do so from anywhere in the world. FVAP administers the Uniformed and Overseas Citizens Absentee Voting Act, or UOCAVA, on behalf of the U.S. secretary of defense and assists voters through partnerships with the U.S. Military, U.S. Department of State, U.S. Department of Justice and election officials from 50 states, U.S. territories and the District of Columbia.

In response to ongoing concerns with the obstacles the military and overseas citizens face when trying to vote absentee, Congress passed the Military and Overseas Voter Empowerment, or MOVE, Act in 2009, granting FVAP the authority to conduct pilot programs to test new technologies that would help with the voting process. In response, FVAP implemented its Electronic Voting Support Wizard—or EVSW–pilot project in 2010, which explored the practical use of technology to streamline the voting process and assist election jurisdictions in meeting their obligations under the MOVE Act to offer an electronic ballot. In its 2013 report to Congress, FVAP stated "the issue of ballot complexity and the ability to exchange data between commercial systems became one of the biggest challenges during project implementation."[1] FVAP has since prioritized the creation of a tool to simplify the data exchange process and also increase the opportunity for UOCAVA-oriented systems to become more viable in the current elections technology marketplace and less likely to fall into obsolescence in the future.

In 2014, FVAP and the National Institute of Standards and Technology, or NIST, engaged in a development effort to create a prototype of such a tool, which yielded the initial Data Migration Tool, or DMT, prototype to address some of the underlying complexities involved with data exchange between systems. One of the key issues is the challenge of importing candidate- and district-level election data into a UOCAVA-tailored system, since many commercial election systems prevent access to certain types of data that are needed to generate an electronic ballot. If data can be exported from these legacy systems at all, the data are typically in an arcane file format that are difficult to decode and then integrate into a more comprehensive collection of useful information.

The DMT allows users, whether they are local election officials or members of stakeholder organizations with a focus on elections, to import data in a specific format—the Voting Information Project, or VIP, specification in either extensible markup language (XML) or comma separated values format[2] —and export the data into the NIST 1500-100 Election Results Common Data Format[3]. The DMT has a web-based interface[4] that allows users with an understanding of elections data and the various input and output formats to more easily work with the system to combine election data into a unified format. The DMT, and the underlying problem it intended to address, represents a starting point for NIST and the elections community. As a research and development effort sponsored by FVAP, the DMT should be seen as a prototype in the truest sense of the word. Based on the development efforts conducted by NIST, lessons learned and the current status of the DMT, the final iteration of the DMT prototype is scheduled to transition into another system that NIST developed called the Materials Data Curation System, or MDCS.

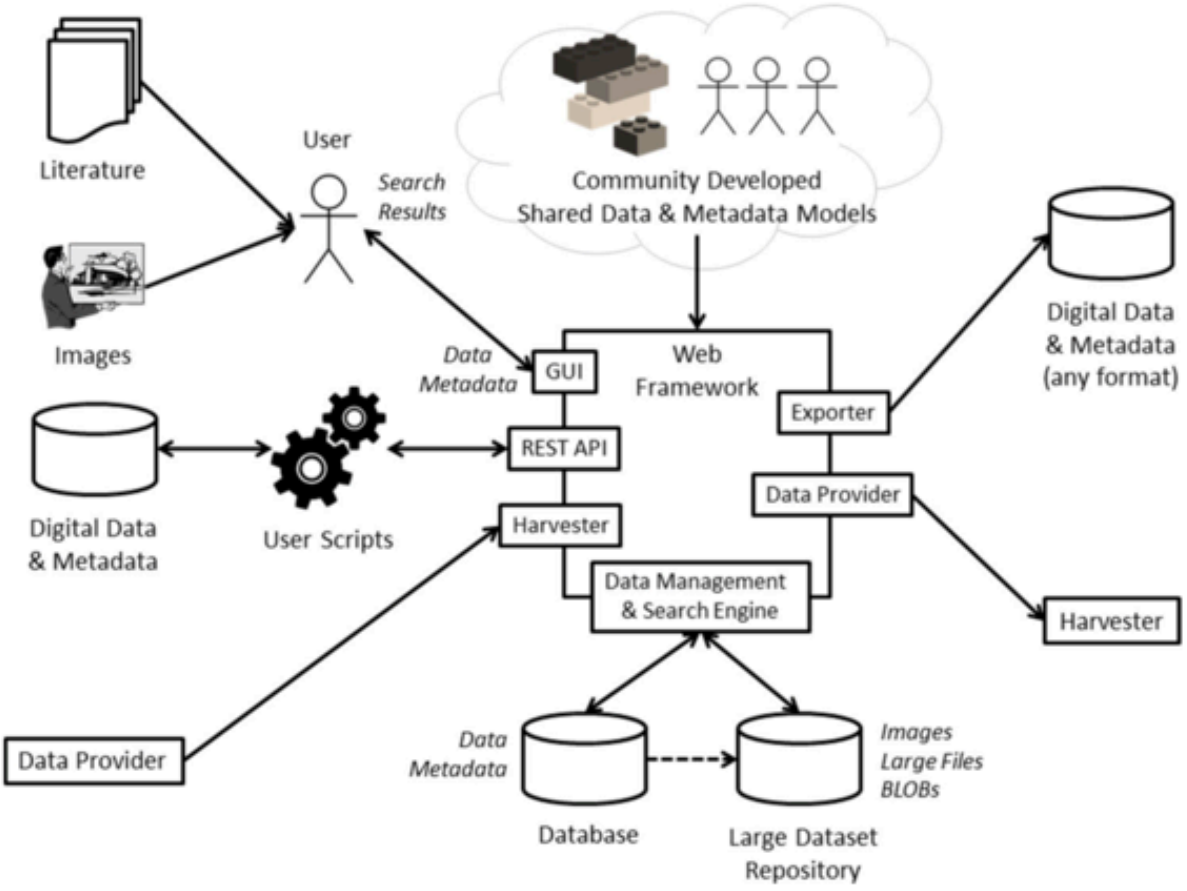
The MDCS is intended to solve the problem of scientific data remaining within academic silos, which parallels many of the same challenges associated with election data across all of the 50 States, territories and the District of Columbia. The goal of the MDCS is to encourage greater collaboration among the scientific community to improve the potential for discoveries and innovation—a goal that can be applied to the election community for the purpose of improving voting for military and overseas voters.

Overall architecture of the Materials Data Curation System[5]

While the DMT and MDCS seem starkly different in nature, they share common characteristics: the desired data are currently difficult to extract because of their isolated nature and the standards for the data are proprietary or loosely defined.

The MDCS has a companion application called the Materials Resource Registry, or MRR. The MRR is a searchable collection of available data and their associated data standards, as well as access and communication protocols. The MRR exists as a resource to help improve exposure and access to the different types of data available within the scientific community.

While there are still substantial obstacles in processing elections data—some of which are addressed in the following scenarios—the process of collecting and standardizing election data, sharing data among state, local and federal agencies, and centralizing data within a state become significantly easier with the use of these tools.



Overall architecture of the Materials Data Curation System[5]

IMMEDIATE VALUES

The DMT laid the framework for working with elections data in the existing technological ecosystem. However, due to the variability of the types of data needing to be transformed, a new system is needed to process and transform existing data into an interoperable format capable of being published and consumed by various systems inside and outside elections agencies. Within the elections industry, there exist only a small number of voting system providers, but a much larger number of ancillary election technology providers of tools such as electronic poll books, election night reporting solutions and election management systems. Moving data between systems or sharing data with outside entities is difficult due to the wide variance of technical offerings governed by few—if any—data standards. Using the MDCS will make this process easier. By defining the data standard in the MDCS or using the MDCS' Application Programming Interface, or API, the MDCS can transform one data standard to another. Should data exist within an Election Management System—or EMS–and a vote tabulation system that would make it possible to generate a specific blank ballot for a voter, and the data are in a format that is unrecognizable to a ballot layout system, the MDCS could be instructed to convert and merge the data files.

While having clear uses in the scientific community, there are ways that the elections community can also leverage the MDCS and MRR. Outlined below are three possibilities for use, their particular obstacles and potential solutions.

STANDARDIZING ELECTION DATA

There are a number of possible uses for data standardization, such as data sharing with other agencies or outside organizations, populating electronic poll books, and creating visualizations for election night results. In each of these scenarios, the media require a specific, rigid set of data, rendering an often-used system export ineffective. In many cases, someone with a background in databases has to generate these specific exports, assuming a resource with these skills is available to the jurisdiction. The aim of the MDCS is to make these tasks significantly more accessible.

OBSTACLES

One of the biggest problems with this effort is that any translation system, including the MDCS, will need minimally standardized data at the outset, which is very similar to the DMT's reliance on a structured dataset available through the efforts made by the Pew Charitable Trusts within the Voting Information Project. Unfortunately, some election systems' exports fail to meet this minimum requirement.

Most Comma/Character Separated Values, or CSV, files resemble a grid: there are a defined number of columns and all subsequent lines of text should have data points that fall into one of the columns. Some tabulation and EMS systems have a non-standardized output file, which have variable number of columns within the same file. Here is one example from Mississippi:

```
0, “GEMS Import Data”, 1, 5, 1, 1, 1, 1
1, “2014 General Election”, “11/4/2014”
2, -1, 1, “United States”
2, -1, 100000275, “State Of Mississippi”
2, -1, 100000047, “US House Of Rep 01”
2, -1, 100000219, “Court Of Appeals 03”
2, -1, 775000024, “Chancery Court 14-2”
2, -1, 775000025, “Chancery Court 14-3”
2, -1, 100000210, “Circuit Court 16”
2, -1, 775002293, “West Point Consolidated School District”
2, -1, 775002294, “West Point Consolidated School District”
3, 0, 3432, “AMERICAN LEGION HUT”
3, 0, 3426, “CAIRO VOTER BUILDING”
3, 0, 3427, “CARADINE VOTER BUILDING”
3, 0, 3428, “CEDAR BLUFF - COMMUNITY CENTER”
3, 0, 3430, “COUNTY COMPLEX”
3, 0, 3453, “HENRY HARRIS BUILDING”
3, 0, 3445, “PHEBA - VOTER BUILDING”
3, 0, 3446, “PINE BLUFF VOTER BUILDING”
3, 0, 3434, “RECREATION CENTER”
```

Along with having no legend to determine what the line length will be or what the columns actually contain, this inconsistency makes the files exceedingly difficult to parse and decipher.

In addition, because of the configurability of many of these election systems, there is no way to guarantee that every export from the same type of system will yield the same export file. While the file above shows some rows that start with the number two having four fields, another system with the same specifications may have the same row type with a greater number of fields.

POTENTIAL SOLUTIONS

While the formatting of the file is a concern in the case listed above, there is a clear pattern: the number of columns potentially changes with the leading number. For every number change in the first column, the data can be divided into separate files.

```
# File 1 – Number of Columns: 8
0, “GEMS Import Data”, 1, 5, 1, 1, 1, 1

# File 2 – Number of Columns: 3
1, “2014 General Election”, “11/4/2014”

# File 3 – Number of Columns: 4
2, -1, 1, “United States”
2, -1, 100000275, “State Of Mississippi”
2, -1, 100000047, “US House Of Rep 01”
2, -1, 100000219, “Court Of Appeals 03”
2, -1, 775000024, “Chancery Court 14-2”
2, -1, 775000025, “Chancery Court 14-3”
2, -1, 100000210, “Circuit Court 16”
2, -1, 775002293, “West Point Consolidated School District”
2, -1, 775002294, “West Point Consolidated School District”
```

Based on the configuration of the system, the type of data in each column for each of the files would still have to be defined, but the MDCS could readily parse each of the separate files.

SHARING DATA

Sharing data is an integral part of an elections office’s operations. State and local agencies need to share data with each other in order to effectively serve the public and perform the work and uphold the responsibilities of their offices. For example, a local tax office and the voter registrar’s office often will share data to determine the precinct of each voter based on residence within a county. There is also sharing of data between the department of motor vehicles, or DMV, office and the voter registrar’s office when a driver registers to vote at the DMV while obtaining or renewing a driver’s license. In many local jurisdictions, there is a voter registrar’s office and an election administration office that must share data to effectively conduct an election.

Federal agencies may also request data of local jurisdictions for compliance reasons, and campaigns and members of the public may ask for data for a variety of reasons. The U.S. Election Assistance Commission, or EAC, collects election data from states and territories in an effort to examine trends and provide guidance and best practices to states. Since the advent of the Help America Vote Act of 2002, the EAC has collected State-by-State information about federal elections every two years. The EAC administers the instrument which collects these data from the states called the Election Administration and Voting Survey, or EAVS.

Academic institutions also collect and examine data, including election data, from a variety of public sources to produce their work. With data analysis tools becoming both more accessible and inexpensive, requests for various types of data have expanded beyond the media and open government organizations to citizens.

The main development goal of the MDCS is to break down data silos between scientific studies in order to catalyze scientific discovery, which can also translate well to the policy research needs of the EAC, FVAP, and the greater academic and election community. To further illustrate the value of the MDCS, consider this example of a potential data sharing scenario. A county elections office needs to share data with the state health and human services, or HHS, agency and the county geographic information systems, or GIS, manager to identify coverage gaps within particular communities. Both entities need access to a collection of data, but some data required by the GIS manager would be unnecessary for HHS and vice versa. Currently, to satisfy both entities, either a local election official–or paid provider–would have to customize an export of data for each entity, or provide the elections data in a potentially nonstandardized format that comes from a system-generated export. The former involves time and money, while the latter may frustrate the receiving agency. If all three agencies had an installation of the MDCS, HHS officials and the GIS manager could find the dataset they need, customize the format and export the data without requiring any direct interaction with the elections office.

OBSTACLES

This scenario is based in a world where MDCS installations are plentiful. Unfortunately, installation of the MDCS is extremely technical in nature and encouraging all agencies of a state to use the software is a herculean task. A large number of local elections offices lack technical resources, making a local installation of the MDCS extraordinarily difficult. Though the MDCS has a graphical user interface to ease the process of managing and transforming data, connecting one MDCS application to another MDCS application—possibly run by another agency—involves knowing highly technical information, such as the internet protocol address of the application, the instance name and server authentication settings.

POTENTIAL SOLUTIONS

While the MDCS may be difficult to grasp for non-technical users, technical users will have access to a significant amount of useful information. The MDCS runs on all three major operating systems—Mac OS X, Windows, and Linux—and all three have detailed instructions for setup and installation. Assuming all necessary prerequisites are fulfilled—programming language, database and other dependencies are all downloaded and installed—a local installation of the MDCS could be up and running in a matter of hours.

NIST developed the MDCS with modern development standards in mind. The API adheres to the Swagger (OpenAPI) specification[6]—a standardized way of documenting and defining an API—making the API easier with which to interact. Since the API allows full access to the capabilities of the MDCS, a developer could use the API to create a script to submit data, which would in turn make it easier to automate the process thereby allowing an election official to bypass some of the more technical challenges of the system.

While both of these solutions still involve a technical user a limition of the MDCS—the potential exists to make the data submission and transformation processes as accessible as possible.

CENTRALIZING DATA

Within the structure of election administration, states represent the most approachable group of stakeholders to drive data centralization through its requirements for establishing a centralized voter registration database under the Help America Vote Act or in their ability to report post-election data to the EAC and FVAP. However, in order to play such a role, states need some view into the local data and a means to transform and transmit data to states in a standardized and repeatable way. The MRR, the companion application to the MDCS, collects the information on all available MDCS instances and allows a single MDCS to connect and receive updates from all relevant MDCS databases. This would allow states to collect, transform and store all datasets published by MDCS installations at a jurisdictional level.

Many decentralized—and even some centralized—states struggle with the process of collecting the data due to the lack of a simple, standardized transport and validation mechanism (i.e. web-based forms require too many validation rules to ensure accuracy and to allow for an off-the-shelf web form service, while using an Excel spreadsheet tends to be difficult for many users). A state could use the MRR as a means of collecting exactly the right data from jurisdiction installations of the MDCS.

This process would assist both states and local jurisdictions during the EAVS data collection process as well as in handling the many requests for data that come from the media, nonprofit groups, universities and other election community stakeholders working to make suggestions for improvements and best practices to the elections process in aggregate.

OBSTACLES

As with many of the aforementioned scenarios, technical comprehension is the biggest hurdle. Similar to the previous data sharing scenario, centralization's success is also dependent on an ecosystem full of MDCS installations, which is difficult to set up without access to a developer. Additionally, jurisdictions often voice concerns about sharing data with a state or sharing responsibility for their voters as the local election officials are the chief administrators of the data and feel personally responsible for the local electorate. The local election officials may view allowing another entity to take over any part of their role as inviting unnecessary risk.

POTENTIAL SOLUTIONS

While many states are decentralized, in States such as Utah and Virginia, local election officials are responsible for the administration of the data, but the database and election software is provided by the state, so it is possible to solve potential local concerns through communication and consensus. While this scenario does not provide a new, statewide election management solution, it does allow for the economies of scale and improved voter communication as outlined above.

SUMMARY

The three scenarios outlined above are only a subset of the vast array of potential uses for election data to showcase the need for standardization. Election officials struggle with collecting their own data as the data is largely unstructured with few constraints to prevent data entry errors. As the current set of data is cleansed, standardized and analyzed, new questions and proposed focus areas will arise about what data should be captured, why it should be captured, and how to capture and structure it. While data standards help the collection of new data, new administrative processes may need to be introduced to capture data necessary to evaluate the efficacy of programs that service particular voting communities.

Based on the original objectives charted by FVAP to provide a mechanism to assist with the translation of data from different sources for use in a UOCAVA blank ballot delivery system, progress has been made as one functional aspect of these objectives is more easily within reach through the use of the DMT and its successors, the MDCS and MRR. Creating a generalized program that converts data from one format—possibly unstandardized and undocumented—to another took a significant work effort and the result is an impressive feat of engineering. The MDCS and MRR are an evolutionary step forward toward the ultimate goal of seamless data migration in the UOCAVA data space, but there is much more work to be done to reach the goal of enabling non-technical election administrators and their staffs to personally undertake this process.

Additionally, obstacles persist due to the lack of a defined UOCAVA data standard, limitations on the overall usability and accessibility of election management systems, and broad adoption of a clearly defined standard to serve UOCAVA blank ballot delivery systems. To prime the election community for success, there is an opportunity for key stakeholders to allocate additional resources toward improving this deficiency by creating easy installation mechanisms, performing usability tests with local election officials, spearheading accessibility tests with members of the accessibility community, and investigating specific use cases that would improve local election administration processes related to data collection.

While local election jurisdictions are accommodating the delivery of blank ballots to UOCAVA voters for the 2016 election, the constraints identified in this use case point to the need for election officials and stakeholders to rely upon the existing patchwork of election data specialists and third party providers to accomplish their objectives. The long-term value proposition of the MDCS and MRR lies within their specific focuses on election data, the promise of the tools to provide the robust infrastructure needed to pursue data standardization and the simplification of data interchanges, and to ensure long-term success with electronic ballot delivery systems and policies derived from sound data.

APPENDIX

DATA MIGRATION TOOL

FVAP and NIST developed the Data Migration Tool, or DMT, as a medium through which data could be transferred between independent election systems. In addition, the DMT could be used to accelerate the adoption of election data standards within jurisdictions prior to election provider integration.

While this tool was in development in 2014, the elections community had witnessed widespread adoption of the Voting Information Project, or VIP, since the project's inception in 2008. A partnership between The Pew Charitable Trusts, Google and the states, the VIP project fostered the adoption of a data standard for publishing non-personally identifiable election information. As VIP continued to integrate data from the states to the local election offices, FVAP and NIST aimed to leverage VIP's success to encourage widespread use of the DMT among election officials.

NIST identified three common scenarios for use of the DMT:

1. A jurisdiction has the necessary formatted data and simply needs to transform it into the NIST 1500-100 Election Results Common Data Format.

2. A jurisdiction has formatted data from two different systems–typically an election management system, or EMS, and voter registration database, or VRDB–and needs to combine them into the 1500-100 format.

3. A jurisdiction needs to manually create the formatted CSV files to transform into 1500-100 format.

For brevity, this walkthrough will only focus on scenario two.

WALKTHROUGH OF THE DATA MIGRATION TOOL

Prerequisites for using the DMT:

- A staff member of the election administration office must have a user account to access the DMT.

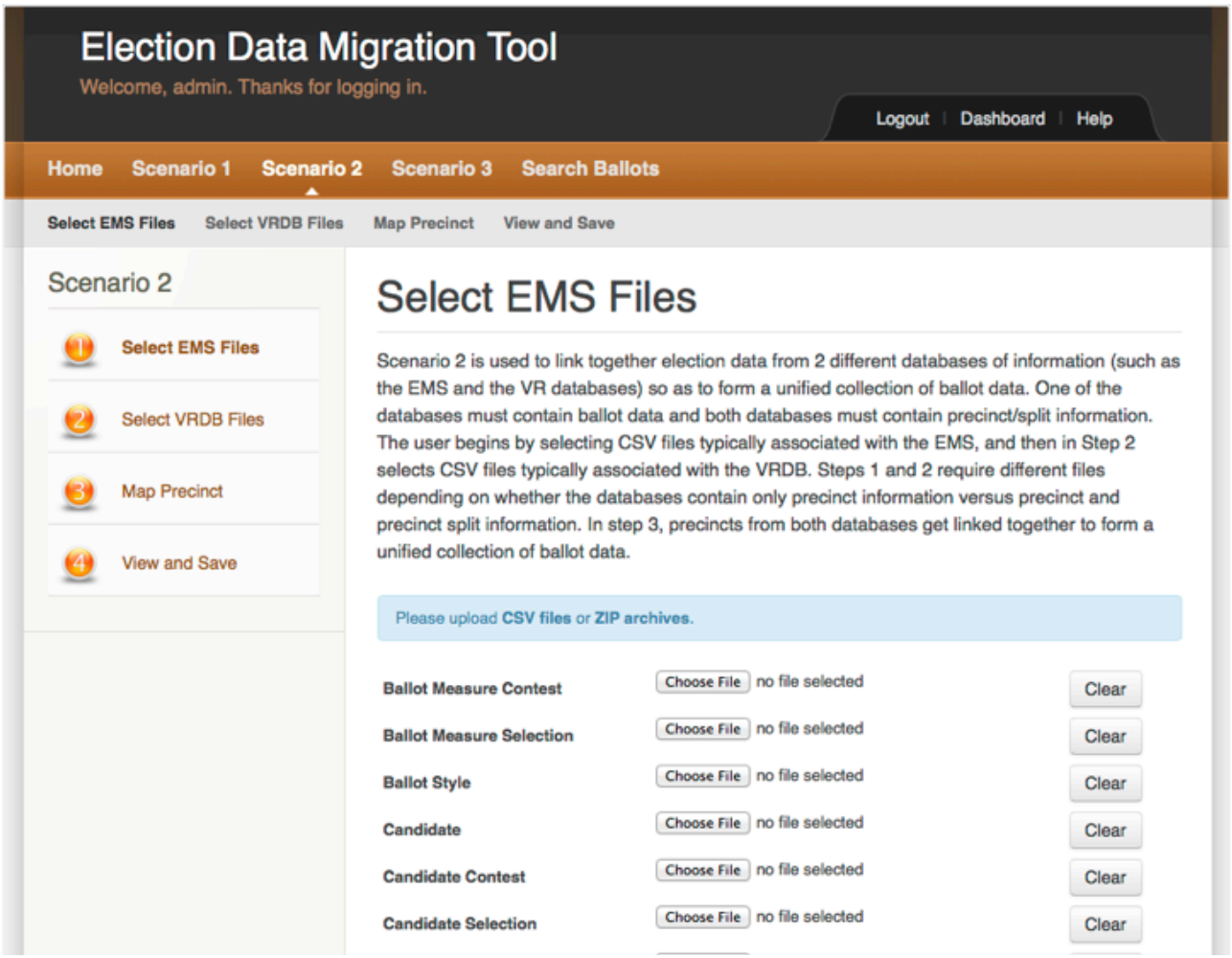
The data for translation must be in VIP CSV specification format,[7] which could be accomplished if the state or jurisdiction either participates in the project or has the technical capability to reformat the existing data. After ensuring that the necessary prerequisites are met, the following steps should be taken:

1. Logging into the tool.

2. Start the process of transforming the two datasets from the EMS and VRDB by clicking "Scenario 2" on the main menu, which will bring the user to the EMS data selection screen (Figure 1).

3. Systematically select each of the appropriate VIP-formatted EMS files by clicking the "Choose

FIGURE 1: EMS DATA SELECTION SCREEN



File" buttons next to each category and selecting the appropriate file from your hard drive. After all the files have been selected, click the "Generate XML File" button at the bottom of the screen, which will generate the first of two NIST 1500-100 XML standard formatted files in this process.

4. Save the generated file to an accessible location on your hard drive and click the sidebar link "Select VRDB Files."

5. On the VRDB data selection screen (Figure 2), the process for providing the data is the same as step 3.

6. Once all the associated files are selected–the result of which can be seen in Figure 3–click the "Generate XML File" button at the bottom of the screen and save the second NIST 1500-100 XML formatted file to an accessible location.

7. After both files are saved, click the "Map Precinct" link in the sidebar.

8. Because the data were exported from two different systems, the two files generated from the previous steps need to be reconciled. First, click on the "Load EMS Names" button.

9. When the "Select an EMS File" appears, click the "Upload a file" button and select the file saved in step 4. Once the file finishes uploading and processing, the precinct names from the

FIGURE 2: VRDB DATA SELECTION SCREEN

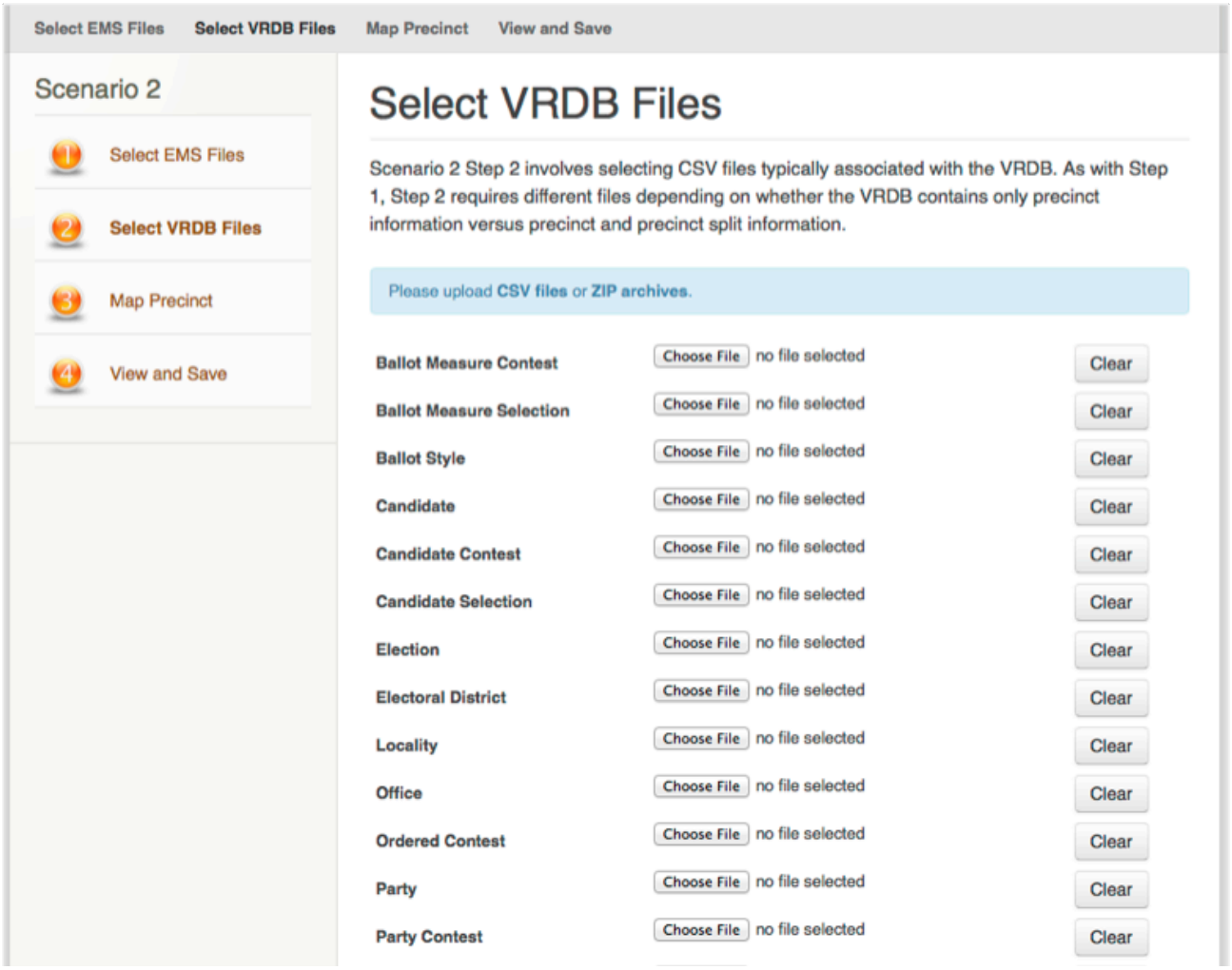


FIGURE 3: SCREEN SHOWING ALL ASSOCIATED FILES SELECTED

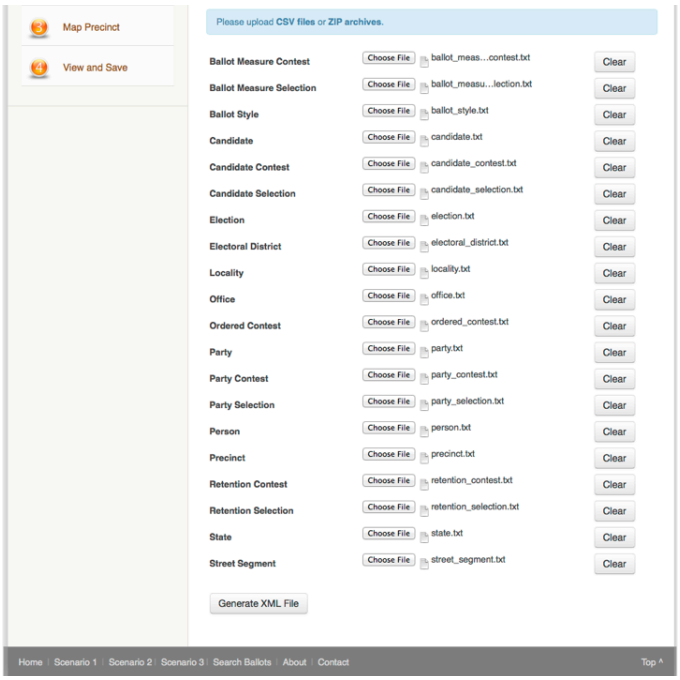
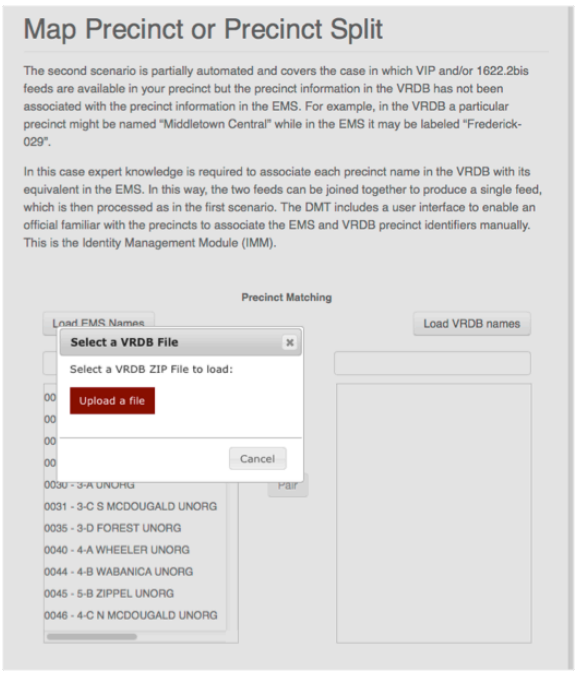


FIGURE 4: UPLOAD XML FILES FOR PRECINCT MATCHING



EMS file will load in the left select box.

10. The process of loading the VRDB names is similar to steps 8 and 9. Click "Load VRDB names" to start the process and upload the VRDB file. Figure 4 shows the screen presented after clicking the "Load VRDB names" button. Behind the "Select a VRDB file" popup box, Figure 4 also shows the EMS file has already been imported.[8]

11. Once both files are loaded, both lists are populated with the names of the precincts that need to be matched.

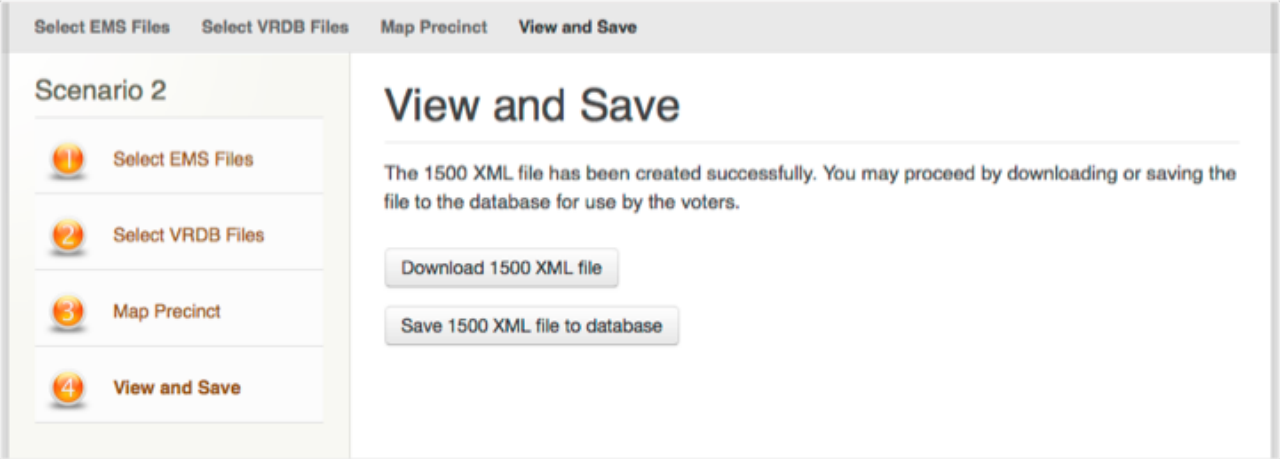
12. After selecting a name in the left column and finding its matching name in the right column, the "Pair" button will remove the names from the two lists and populate the pair in the bottom list. In order to undo a mismatch, select a pair in the bottom table and click the "Unpair" button below (Figure 5).

13. When this process is complete, click the "Export XML" button, which loads the "View and Save" page.

14. The final screen of the process provides the options of downloading the complete 1500-100 file or saving the XML file to the database, which allows for the display of ballot styles by voter address.

FIGURE 5: MATCHING PRECINCTS

FIGURE 6: MATCHING PRECINCTS



FOOTNOTES

[1] Federal Voting Assistance Program, 2010 Electronic Voting Support Wizard Technology Pilot Program Report to Congress, May 2013 (revised July 2013), 13; available from https://www.fvap.gov/uploads/FVAP/Reports/evsw_report.pdf.

[2] Voting Information Project Specification Documentation (accessed August 11, 2016); available from http://vip-specification.readthedocs.io/en/release/.

[3] National Institute of Standards and Technology, NIST Special Publication 1500-100, Election Results Common Data Format Specification,http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-100.pdf.

[4] Data Migration Tool (accessed August 12, 2016); available from https://ballot-test.nist.gov/.

[5] Dima, A., Bhaskarla, S., Becker, C. et al. JOM (2016). doi:10.1007/s11837-016-2000-4, 3.

[6] Open API Specification (accessed July 31, 2016); available from http://swagger.io/specification/.

[7] Voting Information Project Specification CSV Specification Documentation (accessed August 31, 2016); available from http://vip-specification.readthedocs.io/en/latest/csv/index.html.

[8] In a number of jurisdictions, the geopolitical boundaries (e.g. precincts, precinct splits, districts) stored in the VRDB have different identifiers than same geopolitical boundaries stored the EMS. In the best-case scenario, the geopolitical boundaries have the same names and the matching process is simple. In the worst-case scenario, it takes a substantial amount of domain knowledge to match precinct or precinct split codes from both generated lists.

THE COUNCIL OF STATE GOVERNMENTS

ABOUT CSG

Established in 1933, The Council of State Governments is the only organization that takes state government to the next level by creating a multibranch, nonpartisan community. Because of this community, we are able to identify and share best practices and take on the critical challenges of the future in an unparalleled fashion. We conduct research, create forums for innovation and, through our community, ensure the states continue to be recognized as the laboratories of democracy.

CSG OVERVIEW

Governors, legislators, justices, appointed officials and agency directors—our community is composed of officials from all three branches of government from every state and territory in the U.S. Several Canadian provinces also participate in the CSG community through affiliations with CSG regional offices. CSG expertise includes affiliate organizations with specialized knowledge and the CSG Justice Center. Government affairs professionals from Fortune 500 companies, professional associations and nonprofit groups participate in the community through the CSG Associates program.

ABOUT OVI

Many active duty military personnel are located in remote areas abroad and have limited access to state voting information and, in some cases, their ballot. U.S. citizens living overseas also have unique challenges in exercising their right to vote. These challenges are complicated by extreme variation in how states conduct elections and how absentee ballots are processed.

In September 2013, CSG launched a four-year, $3.2 million initiative with the U.S. Department of Defense Federal Voting Assistance Program or FVAP, to improve the return rate of overseas absentee ballots from service members and U.S. citizens abroad.

As part of this effort, CSG's Overseas Voting Initiative maintains two separate advisory working groups. The CSG Policy Working Group is examining military and overseas voting recommendations from the Presidential Commission on Election Administration, as well as other successful programs and practices across the country. The CSG Technology Working Group is exploring issues such as performance metrics and data standardization for incorporation into state and local elections administration policies and practices for overseas ballots. Through the initiative, CSG will provide state policymakers and state and local election officials with best practice guides to ensure the men and women of the U.S. military and Americans living overseas are able to enjoy the same right to vote as citizens living in the United States.

Special Thanks To: Jared Marcotte, CSG Overseas Voting Initiative's Senior Technology Adviser CSG's Overseas Voting Initiative represents a four-year collaboration between CSG and the U.S. Department of Defense's Federal Voting Assistance Program for the purpose of improving the voting process for service members, their families and U.S. citizens living abroad.

CONTACT

Overseas Voting Initiative | (202) 624-3539 | csg.org/ovi | @CSGOverseasVote