What do Databases have to do with Elections?

Stephen Gregory

In the democratic republic that is our United States of America, the future is dependent upon the people, who make their voices heard through voting in elections. This has held true for hundreds of years, but the rise of "big data" lent by the internet and the advancements of computing power and efficiency leads us to an interesting use case of technology for voters: **voter databases**. These voting databases can be used to give election officials invaluable information about the populations of eligible voters and their history of participation in local, state, and federal government. While voter registration and information book-keeping may seem like a trivial task which has little to do with robust data storage at a shallow glance, modern databases prove to be a fundamental part of the complex task that is the management of an entire population of eligible voters, and serve as an interesting case study in the interest of databases in their own right.

Firstly, this topic begs the question: what *is* a voter database? Put simply, a voter database is a comprehensive data storage and retrieval system which contains all of the information about who **is** and **is not** eligible to vote in elections in the U.S., as well as history of participation in elections, income data, work status, and other demographic information [2]. The public interfaces to this kind of database include any of the "Register to vote" websites that one can see in countless Instagram, Facebook, and television advertisements, in addition to the physical voting polls used during elections [3]. When a citizen uses one such website to register to vote, they're using some web application which has been granted access to view and modify a backing voter registration database. Typically, a registering citizen will enter information about their identity, like their name and social security number, along with extraneous information about their job, marital status, residency, and more [1]. This information is then added into a voter database, along with the something on the order of hundreds of thousands or even millions of other voters. These details are insignificant when considering only a handful of voters, but the power of statistics and its applications to large quantities of data allows election officials,

researchers, and more to use this identifying information to gain useful insights on the behaviors of particular demographics of people.

Voter databases vary greatly across state lines, but there are federal regulations which mandate the manner in which each state must maintain its information. In 2002, the *Help America Vote Act* [5] was passed, mandating that each state keep a record of all of its voters, which spurred the first implementations of robust voter databases. Each state is legally required to keep a file of every registered voter in the state, meaning that the unauthorized deletion of registered voters isn't allowed in any state [2]. Beyond this federally mandated bottom line, however, the implementation of databases for each state vary greatly.

One of the issues that must be addressed during this implementation and subsequent maintenance is the management of security and privacy threats. Fair Information Practices (FIPs) form a basis for the privacy policies inherent to the safe management of voter and voter registration information [4]. FIPs address such obviously critical issues as data quality, openness, use limitation, security safeguards, accountability, and more [4]. While not mandated by the federal government, it's important that a given state only collect the minimum viable subset of information needed to keep track of voters, and voters should receive a clear and thorough explanation for the collection of every piece of information in the database. The importance of these principles can not be overstated: if a hacker were to gain full access to a voter database, they would possess the ability to steal the identity of every single registered voter in a particular state. Furthermore, these privacy policies should be visible to the public, so that there is no ambiguity to the registration or voting process for citizens [4]. Security can be achieved through very limited access to the databases, preferably via a small, select number of highly restricted private APIs. Needless to say, there should almost never be direct access to the database itself from anyone other than database administrators, whether the database be implemented using SQL Server database, NoSQL, Oracle, or any other type of database management system (DBMS). Instead, private and public APIs can be developed to utilize all of the useful features of the database, while seriously limiting access to the physical database [6]. Additionally, it could

be very useful to have multiple, up-to-date copies of the database acting as backups, and there might be further utility gained from the creation of separate database backups created with a completely different DBMS altogether (i.e. a SQL Server database in conjunction with a MongoDB and Oracle database). In this way, if a malicious agent exposed a security breach in the publicly-facing primary database, the database administrator for a given state's voter database system could immediately terminate that active database, and set an alternate database instance containing an identical copy of the previous information to be the newly active primary database.

One of the most prevalent subsets of the wide breadth spanned by the term *voter database* is the commonly cited commercial **voter files**. These national databases conglomerate all of the voter registration databases from separate states, and are further supplemented with information from consumer data vendors, private business, credit bureaus, and more [7]. Where the aforementioned voter registration databases aimed to centralize necessary information regarding who is registered and capable of voting in elections, these voter files have a slightly more ambitious target: voter files exist so as to gain detailed, comprehensive information regarding every citizen in the United States, including those who are not registered to vote. Where voter registration files and other federally mandated database records aim to maintain the very smallest amount of information necessary to keep track of citizen voting, voter files do not share the same goals [7]. Because these commercial voter files exist in the free market, they inherently exist not as a vehicle to leverage regulation and public welfare, but instead as a valuable resource to be treated as any other in a capitalist economy: free to be sold, traded, priced, and held by anyone with access. A recent study into the coverage of these voter files over the entire U.S. population shows interesting results: in one study by the Pew Research Center, an average of 69% of citizens taken from a nationally representative survey panel of U.S. adults were mapped to one of a choice of five commercial voter files [7]. This is an important discovery, leading one to believe that a potential two-thirds of the adult U.S. population may be comprehensively tracked in at least one of these commercial voter files. This information may prove to be crucial in elections; the influence of social media platforms and effectiveness of advertisements using such platforms

is difficult to deny, and individuals' information can be bought and sold to political parties and politicians so as to target advertisements to the perfect demographics of people. These kinds of targeted advertisements wouldn't exist without voter files, and the right advertisements targeted to the right audience can and frequently will be the greatest factor in the outcome of any given marketing campaign. Why should elections be any different?

Elections in the U.S. provide a prototypical example of an ideal use case for databases. Hundreds of millions of citizens may potentially participate in elections, and without a robust solution for the storage of all of these individuals' data, there is simply a huge amount of intelligence and insight that is being left on the table. If we were to forego the use of voter databases, a few things would happen. There would be many fewer people participating in elections, advertisements would be much less effective, wasting valuable time and resources for political candidates, and there would be no way to maintain histories of whether or not citizens were registered to vote, or even if they had already voted in a particular election! Therefore, it should seem obvious that voter databases are absolutely crucial to the function of a modern republican democracy. However, we should also be wary of the dangers of massive amounts of personal data being stored by the free market *and* the government, and it's important in our design of databases that we use good principles such as redundancy and secure design at the systems-level. Furthermore, it's important that privacy agreements and policies be clearly, effectively, repeatedly communicated to citizens. The future of voter databases is exciting and bright, and will rely on the contributions of politicians, database administrators, data scientists, and countless others to be fully realized in a powerful, safe, and secure manner.

References

[1] “Access To and Use Of Voter Registration Lists,” National Conference of State Legislatures, Aug. 5, 2019. [Online]. Available: https://www.ncsl.org/research/elections-and-campaigns/access-to-and-use-of-voter-registration-lists.aspx

[2] D. Desilver, “Q&A: The growing use of ‘voter files’ in studying the U.S. electorate,” Pew Research Center, Feb. 15, 2018. [Online]. Available: https://www.pewresearch.org/fact-tank/2018/02/15/voter-files-study-qa/

[3] “Online Voter Registration,” National Conference of State Legislatures, Aug. 19, 2020. [Online]. Available: https://www.ncsl.org/research/elections-and-campaigns/electronic-or-online-voter-registration.aspx

[4] P. Hawthorn, B. Simons, S. M. Bellovin, C. Clifton, L. Coney, R. Gellman, H. Hochheiser, R. S. Poore, A. Rosenthal, D. Wagner, R. N. Wright, “Statewide Databases of Registered Voters: Study Of Accuracy, Privacy, Usability, Security, and Reliability Issues commissioned by the U.S. Public Policy Committee of the Association for Computing Machinery”, *Association for Computing Machinery Advancing Computing as a Science & Profession*, p. 1+, Feb. 2006. [Online]. Available: https://www.acm.org/binaries/content/assets/public-policy/usacm/

[5] The United States Department of Justice, “The Help America Vote Act of 2002”, *The United States Department of Justice*, Aug. 24, 2020. [Online]. Available: https://www.justice.gov/crt/help-america-vote-act-2002

[6] U. Sivarajah, M. M. Kamal, Z. Irani, V. Weerakkody, “Critical analysis of Big Data challenges and analytical methods,” *Journal of Business Research*, vol. 70, p. 263-286, Jan. 2017. [Online serial]. Available: https://www.sciencedirect.com/journal/journal-of-business-research/vol/70/suppl/C

[7] R. Igielnik, S. Keeter, C. Kennedy, B. Spahn, “Commercial Voter Files and the Study of U.S. Politics”, Pew Research Center, Feb. 15, 2018. [Online]. Available: https://www.pewresearch.org/methods/2018/02/15/commercial-voter-files-and-the-study-of-u-s-politics/