

# **2006 Forensic Election Data Collection and Analysis**

To be conducted by Campaign Scientific, Stephanie Singer, Director

**September 2006**

## **Proposal Overview**

We propose a comprehensive election-data collection and analysis effort in selected jurisdictions for the General Election of 2006. This would be a component of a pilot project to verify the accuracy and integrity of electronic voting technology, and to prepare for a full national 50 state election vote-count protection effort for the 2008 elections.

### **Contents**

1. What is Election Data?
2. Does the US Need Another Election Data Collection Effort?
3. Forensic Data Analysis
4. Data to Be Collected
5. Key Steps in Data Collection and Analysis
6. Budget

## **1. What is Election Data?**

To determine who won an election requires knowing only the total number of votes earned by each candidate in the jurisdiction; to monitor an election requires more detailed data. Data must be broken down by precinct and voting method. For each subdivision there is a total number of votes for each candidate. These numbers of votes, for every possible combination of precinct, voting method and candidate, constitute the best database of “vote counts”. In addition, there is important supplementary data to be collected for each precinct. Finally, the data may change as it is processed, i.e., as the information travels from machine to precinct captain to central counting office computers to officially reported results. Because of the power of computers to analyze and consolidate data, the general strategy is to collect as much, and as finely-grained, data as possible.

The precinct is the smallest demographic unit for which election results are available. It is not possible to collect election results on a smaller scale (though it may be possible to track individual machines within one precinct). On the other hand, problems can occur on as small a scale as an individual precinct. For example, one of the hypotheses consistent with the available

data for the 2004 election is that official vote totals for Bush were increased selectively in precincts where voters voted heavily, but not overwhelmingly, for Bush.

Even within a single precinct, the vote count can be broken down by voting method. Depending on the state, voters can vote by regular ballot at the polls, by absentee ballot, by provisional ballot or by early ballot. Regular ballot tallies can be broken down further by the machine used. A detailed breakdown of election results by vote type can catch suspicious discrepancies that would otherwise go unnoticed. For example, an increase in the reported total number of absentee ballots could mask a large undervote due to machine error.<sup>1</sup>

Aside from the vote-count numbers, supplementary information about the precinct is helpful to any large-scale analysis. At a minimum, one should collect the number of eligible voters in each party, as well as the number of voters who voted by each method. Furthermore, demographic data is useful, as is historical data. Any available exit-poll data should be included as well.

Finally, in the course of a single election, vote counts are transmitted and processed in various ways. Precinct-level machine totals are sent to the county Board of Elections for aggregation. *A priori*, every stage of the process is vulnerable to error and manipulation. Fortunately, because the vote-counting is (at least in theory) open to scrutiny, there are opportunities to collect data at several points during the process.

All this data (precinct-level vote counts broken out by machine type monitored for changes during the time between the close of polls and the certification of the election, and combined with relevant demographic and historical data) can be combined into a powerful database ripe for the identification both of inconsistencies demanding explanation and statistical trends deserving of further investigation.

## 2. Does the US Need Another Election Data Collection Effort?

Yes. Steve Freeman and Ken Warren have made the case elsewhere<sup>2</sup> that our elections are in need of monitoring. While there are various institutions collecting election data in the United States, none are collecting the data necessary for comprehensive nonpartisan monitoring of election machinery. Political parties, some academic institutions, and government agencies do collect data, but the availability and reliability of this data depends on their individual motives and funding levels. In any case, the data they collect is exclusively official data, collected at the end of the vote-tallying process. Such data is not sufficient for aggressive monitoring of the vote count. The data collection initiative we propose here is necessary to any serious effort to monitor and audit elections.

Political parties, at both the national and local levels, collect extensive amounts of data. However, they have no incentive to share this data. Worse yet, they may have incentives to falsify any data they release. For our purposes, this data is probably unavailable and surely unreliable.

Some academic institutions sometimes collect election-related data, but this data tends to be spotty. For example, the Pennsylvania State Data Center at Penn State (<http://pasdc.hbg.psu.edu>) has some historical registration data for the state, but only for a period in the 1990's. Other sites (such as <http://uselectionatlas.org>) have statewide aggregate data, but not precinct-level data.

State government records of election results vary in quality from state to state. For example, in New Mexico, sufficiently detailed election results are available in electronic format, while in Pennsylvania they are not.

Even if these institutions collected data reliably and made it freely available, it would not suffice for the purposes of monitoring the vote-counting. To monitor the count, one must collect not only the official results, but various intermediate results. This kind of comprehensive data collection should be done by independent, non-partisan watchdog groups. The National Election Data Archive (<http://electionarchive.org>) is attempting to collect this data nationwide on a volunteer basis but, as the scope of this proposal should make clear, volunteer efforts will not suffice for this project. Thorough, strategic data collection is a crucial component of any attempt to monitor an election or the performance of vote-counting machinery.

### 3. Forensic Data Analysis

It is not enough to collect data; one must analyze it as well. One task is to compare similar data from several sources. A second is to check for logical consistency of the data as a whole. A third is to apply the most sophisticated mathematical and statistical techniques for identifying patterns in the data.

Because we are collecting overlapping data from several sources, we will want to check that the sources agree. For example, we will have two sources for the machine vote-count in some precincts: machine tapes collected at the precinct and readouts from the machine cartridges collected at the Board of Elections. In an ideal world, these vote-counts match. We must determine whether they match in the jurisdictions we have targeted.

Because we are collecting many different kinds of data, we will want to check that the data are logically consistent. For example, in each precinct we can compare the reported number of voters with the total reported number of votes in each race. In an ideal world, the first number is equal to or (because a voter can choose not to vote in any particular race) larger than each of the other numbers. Past US elections have been less than ideal in this regard. For example, in Franklin County, Ohio, in the 2004 general election an electronic voting machine gave one candidate 3,893 extra votes in a precinct where only 638 voters cast ballots.<sup>3</sup> Checking for logical consistency is crucial to election monitoring.

Because we are collecting a large amount of data, we will want to look for trends in the data. There are now sophisticated algorithms for pattern detection, known colloquially as “data mining”. Data mining allows supermarkets to find needles (e.g., the fact that men who buy diapers on Thursdays also tend to buy beer) in the haystack of a database of all purchases by all shoppers who carry discount tags on their key chains. Data mining can allow us to find

unsuspected patterns in our database of election data. If there is systematic error or fraud in the election, it has to leave a trace somewhere. Without data mining, we can look for the trace only where we can guess it might be (e.g., correlated with machine type). With data mining, we can find it in places we might not have thought of.

These analyses will provide a much-needed audit of the election procedures and machinery. In addition, they will generate greater understanding of the actual trends in election data, which will aid future election protection efforts.

#### **4. Data to be Collected**

- Voter file used to produce pollbooks
- Precinct-level machine vote-counts at the polling place at close of polls
- Number of absentee, provisional, etc., voters in precinct according to Board of Elections
- Precinct-level machine vote-counts at central tabulation location
- Precinct-level absentee, provisional, etc., vote-counts at central tabulation location
- “Unofficial” election results reported by the media
- Official, certified election results from county Boards of Election
- Census data, especially precinct population
- Historical election data from state and county records
- Historical election data from academic and other non-profit sources
- Exit poll data.

#### **5. Key Steps in Data Collection and Analysis**

The timeline for the bulk of the proposal is September 2006 through January 2007. There may be dissemination activities after that period.

1. Get started: Meet with the core collection & analysis team, develop overall game plan
2. Hire key personnel and set up office
3. Choose counties and precincts within counties
4. Recruit and train data collectors
5. Obtain preliminary data
6. Research Election-Day and post-Election-Day procedures in targeted jurisdictions
7. Purchase essential equipment (e.g., software and computers)
8. Ensure access to polling place and location of central counting
9. Collect supplementary and historical data

10. Prepare to receive and process data quickly
11. Collect vote counts after close of polls
12. Perform fast analyses
13. Perform comprehensive analyses
14. Collect additional data as necessary
15. Analyze results and draft report
16. Distribute findings to appropriate groups and interested parties

## **6. Budget**

### **6. 1. Data collection and analysis management team**

The team will be responsible for designing and overseeing the entire data collection project. Stephanie Frank Singer, director of the Pennsylvania Election Transparency Project, has experience collecting election data from state and county agencies all over Pennsylvania. Another manager with complementary experience (either geographically or managerially) will be hired. We will also hire an office manager to help with organization and accounting.

### **6. 2. Programming and Analysis**

- Design and implement database architecture.
- Design and implement consistency checks and data mining routines, some fast and some comprehensive.
- Analyze the collected data, synthesize results and create new routines as suggested by findings.

### **6. 3. Consultants**

Legal consultants will help us navigate laws pertaining to election oversight and data collection so that we can collect as aggressively and as accurately as possible. Scientific consultants will help us fashion the best possible database architecture and analysis algorithms.

### **6. 4. Data**

There are costs associated with collecting data, even from government and other public databases. Databases collected by private firms (such as the useful historical data from the non-profit organization National Committee for an Effective Congress) are priced at market rate.

### **6. 5. Hiring and Training Data Collectors**

Training per se will not require additional money because it will be done by the management team and staff. Personnel from the management team and coordinator team can take responsibility for recruiting and training interviewers, as part of their job description, but data collectors have to be paid to attend session(s). We will recruit and train 20% more than anticipated Election Day staff necessary to account for attrition.

## **6. 6. Rent, Hardware and Software**

- Office space with phone and internet for management team.
- Office equipment for management and data consolidation.
- Laptops for Election Day data collection. We assume that most, but not all, of the data collectors will have laptops of their own that they can use.

## **6. 7. Data Collection**

- Stipends for data collectors, one for each county. The rate is high in order to attract technically competent data collectors who also know how to press for access to the data.
- Travel and lodging for some of the data collectors. In the larger counties we can expect to find data collectors who live near the Board of Elections.

## **6.8. Data Entry**

- Data entry specialists responsible for organizing and entering the various data sets and merging the data.
- Data verification from field to central headquarters. Final verification would be made by data entry specialists and the management team at central headquarters.

## **6.9. Travel**

- In addition to the foreseeable travel for the purposes of training and data collection, managers or experts may have to travel to the site of a contested election or an uncooperative Board of Elections.

## **6.10. Dissemination of Results**

- Production and distribution of written reports.
- Travel to speaking opportunities such as the February 2006 symposium on vote-counting at the AAAS meeting in San Francisco.

## **Endnotes:**

---

<sup>1</sup> Dopp, Kathy, "How Could Election Officials Ensure Accurate Vote Counts?" (downloaded on 30 July 2006 from [http://www.electionarchive.org/ucvAnalysis/US/election\\_officials/Audits\\_Monitoring.pdf](http://www.electionarchive.org/ucvAnalysis/US/election_officials/Audits_Monitoring.pdf)).

<sup>2</sup> Freeman, Steve and Ken Warren, "2006 Election Verification Exit Poll Proposal Prepared for People for the American Way, 4 July 2006.

---

<sup>3</sup> Associated Press, "Computer error at voting machine gives Bush 3,893 extra votes", 5 November 2004 (downloaded on 31 July 2006 from <http://www.votersunite.org/article.asp?id=3702>).