06-CV-00726-RCPT

UNITED STATES DISTRICT COURT
WESTERN DISTRICT OF WASHINGTON

WASHINGTON ASSOCIATION OF CHURCHES, as an organization and representative of its members; WASHINGTON ASSOCIATION OF COMMUNITY ORGANIZATIONS FOR REFORM NOW (ACORN), as an organization and representative of its members; ORGANIZATION OF CHINESE-AMERICANS – GREATER SEATTLE CHAPTER, as an organization and representative of its members; CHINESE INFORMATION & SERVICE CENTER, as an organization and representative of its clients; FILIPINO AMERICAN POLITICAL ACTION GROUP OF WASHINGTON, as an organization and representative of its members; KOREAN AMERICAN VOTERS ALLIANCE, as an organization and representative of its members; SERVICE EMPLOYEES INTERNATIONAL UNION (SEIU) – LOCAL 775, as an organization and representative of its members; and WASHINGTON CITIZEN ACTION, as an organization and representative of its members,

Plaintiffs,

vs.

SAM REED, in his official capacity as Secretary of State for the State of Washington,

Defendant.

Civil No. 06-0726

DECLARATION OF ANDREW BORTHWICK

IN SUPPORT OF PLAINTIFFS' MOTION FOR A PRELIMINARY INJUNCTION

HILLIS CLARK MARTIN & PETERSON, P.S.

500 Galland Building, 1221 Second Ave
Seattle WA 98101-2925
206.623.1745; fax 206.623.7789

Pursuant to 28 U.S.C. § 1746, I, Andrew Borthwick, hereby declare as follows:

1. I am the President and Chief Executive Officer of ChoiceMaker Technologies, Inc. ("ChoiceMaker"), which I co-founded in 1998. ChoiceMaker is a data quality company specializing in the design and development of record-matching software. I submit this declaration in support of Plaintiffs' Motion for a Preliminary Injunction.

Background

2. I earned a B.A. from Oberlin College in 1988, graduating Phi Beta Kappa. I earned an M.S. and Ph.D. in Computer Science from New York University. My Ph.D. was awarded in 1999. My doctoral dissertation discussed a maximum entropy approach to named entity recognition, which in broad terms involves a learning technology that builds a model of the human decision-making process for identifying and categorizing proper names, in order to find names in context in newspaper text. For example, my dissertation discussed an approach useful for distinguishing articles about Calvin Klein, the individual, from articles about Calvin Klein, the company. I co-founded ChoiceMaker to apply the technology discussed in my dissertation to the record-matching field. My office is located at 48 Wall Street, 11th Floor, New York, New York 10005.

3. My academic expertise is in the fields of record-matching, machine learning, and computational linguistics. Of greatest relevance here is my background in the first, "record matching," which is a common term of art in the field of information science (also referred to as "informatics"). Record-matching refers to the process of identifying entries in a database (known as "records") that pertain to the same entity or person represented in other records, either in the same database or in another database. Often, these records are matched by comparing particular categories of data (known as "fields") within each record, such as a person's first name, last name, and date of birth. The science and challenge of record-matching involves identifying matching entries in the face of errors in one or both sets of data, or of inconsistencies between the data. I was awarded U.S. Patent No. 6,523,019 in 2003 for a machine learning approach to record-matching. I also am the co-inventor of two other processes

HILLIS CLARK MARTIN & PETERSON, P.S.
500 Galland Building, 1221 Second Ave
Seattle WA 98101-2925
206.623.1745; fax 206.623.7789

concerning record-matching with patent applications pending before the U.S. Patent and Trademark Office.

4. In 1998, I founded ChoiceMaker, and I was joined as co-founder by my business partner, Arthur Goldberg. ChoiceMaker is a data quality company specializing in record-matching software. Among other functions, ChoiceMaker software allows entities to identify corresponding records within and between databases. From 2000 to 2005, ChoiceMaker won a prestigious series of Small Business Innovation Research grants from the National Science Foundation for its research into a maximum entropy approach to approximate record matching.

5. In 2005, ChoiceMaker was awarded a contract with the U.S. Centers for Disease Control ("CDC") to use our record-matching system for the National Electronic Disease Surveillance System. This CDC surveillance project tracks the emergence and incidence of diseases in all 50 states in order to facilitate reporting to the CDC and to identify outbreaks of infectious diseases. The ChoiceMaker system was also purchased by nine states to track the academic records of every student in K through 12 public education, to support implementation of the No Child Left Behind Act. ChoiceMaker has also won multiple contracts with the New York City Department of Health to track immunizations of children, lead tests, and the incidence of communicable diseases – which requires matching records from laboratories, hospitals, and clinics, in the face of numerous errors. ChoiceMaker was also used for the World Trade Center Health Registry to compile a register of everyone near the World Trade Center on 9/11 in order to track long-term health issues. ChoiceMaker was also awarded a contract with the New South Wales, Australia, Department of Health to develop a system for epidemiological research.[1]

6. I have published on, among other things, techniques involved in record-matching, including articles for peer-reviewed conferences in 1999 and 2004. A complete list of my publications is included in my *curriculum vitae*, a copy of which is attached as Exhibit A. I was invited to speak on record-matching at the First Workshop on Data

[1] ChoiceMaker has never been, and is not now, affiliated with ChoicePoint, Inc., Database Technologies (DBT Online), or any of their successors in interest.

HILLIS CLARK MARTIN & PETERSON, P.S.

500 Galland Building, 1221 Second Ave
Seattle WA 98101-2925
206.623.1745; fax 206.623.7789

Cleaning, Record Linkage, and Object Consolidation, in conjunction with the Association for Computing Machinery Special Interest Group on Knowledge Discovery and Data Mining's Ninth International Conference on Knowledge Discovery and Data Mining, in Washington, D.C., on July 17, 2003. In addition, I have given presentations at the Massachusetts Information of Technology's annual International Conference on Information Quality, the annual Information Quality Conference, and the annual National Immunization Conference, among others.

7. I regularly review publications in the record-matching field, in order to ensure that I am well versed in the current state of the art. Representative publications of this sort include *Institute of Electrical and Electronics Engineers ("IEEE") Transactions on Pattern Analysis and Machine Intelligence* and *IEEE Transactions on Knowledge and Data Engineering*. I am a member of the IEEE, the world's leading professional association for the advancement of technology, and the Association for Computing Machinery, an international scientific and educational organization dedicated to advancing the arts, sciences, and applications of information technology. I have not previously given testimony as an expert witness. My billing rate for this matter is $150 per hour.

Summary of Conclusions

8. For this case, I was asked to examine whether the record-matching protocols used in Washington State's new voter registration program will result in a significant number of errors. In particular, I was asked to explain whether and why, in my professional opinion, errors endemic to information gathering, entry and maintenance, along with immaterial differences and inconsistencies across different databases, will result in the failure to match information from two different sources pertaining to the same individual. In the technical language of my field, the question I explored was whether an exact comparison of the characters in multiple fields of two different records, each filled out (or "populated") at different times by manual data entry from different handwritten forms or from information given orally to a data entry clerk, would likely result in substantial numbers of "false negatives" – that is, registration records for which the name and identifying information do not "match" the name and

HILLIS CLARK MARTIN & PETERSON, P.S.
500 Galland Building, 1221 Second Ave
Seattle WA 98101-2925
206.623.1745; fax 206.623.7789

information in another database when, in fact, both records reflect the same person. For example, a "false negative" will result when a woman registers to vote with her married name, her maiden name is listed on her Social Security record, and the two records fail to "match."

9. To prepare to give my opinions in this case, I read the applicable federal and State laws and regulations, including relevant portions of the Help America Vote Act (and in particular 42 U.S.C. § 15483) and RCW 29A.08.107, my own academic and professional work, and relevant publications in the field. Those publications include papers in the *Statistical Research Report* series of the U.S. Bureau of the Census, articles published in peer reviewed journals such as *Computers and Biomedical Research* (now known as the *Journal of Biomedical Informatics*) and the *Journal of the American Medical Informatics Association* (AMIA), and proceedings of the AMIA's annual symposium.

10. To learn about how the record-matching process in Washington works, I also reviewed documents provided to the Plaintiffs by Washington election officials pertaining to the matching process, including the *"Social Security Verification" System Specification* distributed by the American Association of Motor Vehicle Administrators in August 2004. In particular, I reviewed the Help America Vote Verification ("HAVV") process (also referred to as a "transaction") and the matching protocol described on pages 26 and 27 of that document; a copy of the relevant pages is attached as Exhibit B. I also reviewed the "Voter Registration Identity Verification Procedures" posted on the website of the Washington Secretary of State, and attached as Exhibit C. I also reviewed the public presentation of Peter Monaghan, the Social Security Administration's Senior Advisor to the Office of Programs, dated February 6, 2006, relating to a recent audit the Social Security Administration has performed relating to its matching of voter registration information. A copy is attached as Exhibit D. I reviewed similar reports of audits of the voter registration information matching process by election officials in New York City. A copy is attached as Exhibit E.

11. Based on my academic and professional experience with record matching, and additional research and analysis I performed, and based on my understanding of Washington's matching protocol, I conclude that Washington's voter registration matching

HILLIS CLARK MARTIN & PETERSON, P.S.
500 Galland Building, 1221 Second Ave
Seattle WA 98101-2925
206.623.1745; fax 206.623.7789

protocol will result in a significant number of "false negatives": records that pertain to the same individual but which are unable to be matched. If Washington conditions registration on a successful "match," a significant number of valid voter registration applications will be rejected as a result. These errors are likely to occur even if the applicants do not make any mistakes or provide any incorrect information on their registration forms. It also is my opinion, based on studies of similar protocols, that the rate of such "false negative" errors will be substantial, and could reach as high as 30%.

The Process of Record Matching in Washington

12. I understand that at the beginning of this year, the Secretary of State of Washington, in conjunction with other State and federal agencies, began to attempt to match certain information provided on new voter registration forms with information stored on other databases. Based on a document provided by Washington election officials pertaining to the matching process, a copy of which is attached as Exhibit F, I outline in broad terms below my understanding of the process of record matching as it is performed in Washington.

13. *First*, I understand that citizens fill out a voter registration form by hand with their identifying information. That information includes: name, date of birth, and either a driver's license number or the last four digits of their Social Security number (if the applicant has such a number). A true and correct copy of Washington's registration form available online from the website of Washington's Secretary of State is attached hereto as Exhibit G. Registrants may also supply their identifying information orally to a data entry clerk. The completed forms are then submitted to State or county officials.

14. *Second*, I understand that data entry operators working for the State or county will input the data contained on the voter registration forms into one or more databases that serve as temporary, electronic storage for such new registration records.

15. *Third*, I understand that each new electronic registration record will be submitted to State officials, who will cause certain pieces of information in the registration record to be compared automatically either to the Social Security Administration database or to

HILLIS CLARK MARTIN & PETERSON, P.S.

500 Galland Building, 1221 Second Ave
Seattle WA 98101-2925
206.623.1745; fax 206.623.7789

the State Department of Licensing database. This is done in an attempt to "match" the information contained in the voter registration record to the information contained in the database. For matching with the Social Security Administration database, my understanding is that Washington is using a protocol in which each character of the first name, each character of the last name, each character of the year of birth, each character of the month of birth, and each character of the last four digits of the applicant's Social Security number, as entered in the voter registration record, must match exactly with the corresponding character or the corresponding field of the Social Security Administration database. For matching with the Department of Licensing database, my understanding is that Washington is using a protocol in which at least the driver's license or non-driver's identification card number must match exactly with the characters of the corresponding field in the Department of Licensing database.

16. *Fourth*, I understand that if the State finds a "match," the person who filled out the form, if otherwise eligible, will be registered to vote. If the State does not find a "match," the person who filled out the application will not be registered to vote, although I understand that State officials will attempt to correspond with such applicants to try to resolve the problem.

17. Based on this information, it is my opinion that record matching as I understand it will be conducted in Washington will likely result in high numbers of false negatives. That is, I am confident that attempts to match information in new voter registration records to information in other State and federal databases will fail for reasons unrelated to the accuracy of information provided by the applicants.

Common Errors Related to Record Matching

18. There are several reasons why large databases are prone to errors that make the process of record matching imperfect. The point is a rather basic one, but it has profound consequences when attempting to match individual records in one large database with records in another database: typos, misplaced information, incorrectly transcribed data, and

HILLIS CLARK MARTIN & PETERSON, P.S.
500 Galland Building, 1221 Second Ave
Seattle WA 98101-2925
206.623.1745; fax 206.623.7789

immaterial spelling and punctuation differences in either one or both of the databases may result in two records for the same person not "matching."

19. **Data Submission.** Errors within individual records of large databases may be caused by mistakes in data submission. I am not referring to false information, but mistakes in the form in which the data is submitted. Such mistakes can include minor errors made by individuals filling out forms, such as writing information in one place when the information should be written in another. These immaterial mistakes may appear in the registration record or in the government database being matched – or in both. For example, a person may write her day of birth in a space reserved for the month of birth. If this were to occur in Washington, that person's birth date as entered from her registration record will not exactly match the birth date as recorded on the database with which the State will compare her identifying information.

20. **Data Entry.** Errors within individual records of large databases may also be caused by mistakes in the process of entering the data in the computer. Such errors may occur when an operator strikes an incorrect key, incorrectly hears information given orally, or incorrectly reads information from a form. For example, a data entry operator may type an "a" when an "o" is written, or type a "d" when a "c" and an "l" are written together. Common data entry errors also include:

- omitting characters (*e.g.*, "THOMAS" becomes "TOMAS," "JOHN" becomes "JON");
- adding characters (*e.g.*, "THOMAS" becomes "THOMMAS," "OWEN" becomes "OWENS");
- transposing characters (*e.g.*, "THOMAS" becomes "TOHMAS," "KREIDLER" becomes "KRIEDLER");
- substituting characters (*e.g.*, "THOMAS" becomes "THOMAS" or "THIMAS," "REID" becomes "REED"); or
- any combination of the above.

HILLIS CLARK MARTIN & PETERSON, P.S.

500 Galland Building, 1221 Second Ave
Seattle WA 98101-2925
206.623.1745; fax 206.623.7789

21. Other data entry errors may occur when an operator enters the information in the wrong field (*e.g.*, inverts the day and month in the fields provided for the date of birth). Operators also separate compound last names into the "middle name" and "last name" fields or, conversely, combine a middle and last name into a single last name (*e.g.*, "WILLIAM" "BUTLER" "YEATS" becomes "WILLIAM" "BUTLER YEATS"). Such errors include:

- omitting fields (*e.g.*, "JAMES THOMAS" becomes "THOMAS");
- adding fields (*e.g.*, "JAMES THOMAS" becomes "JAMES J THOMAS" or "MR JAMES THOMAS" or "CAPT JAMES THOMAS");
- transposing fields (*e.g.*, "JAMES THOMAS" becomes "THOMAS JAMES", or "LU BAO" becomes "BAO LU");
- substituting fields (*e.g.*, "JIMMY THOMAS" becomes "JAMES THOMAS");
- improperly separating fields (*e.g.*, "JAMES" "THOMAS-SMITH" becomes "JAMES" "THOMAS" "SMITH");
- improperly combining fields (*e.g.*, "MARY" "ANN" "THOMAS" becomes "MARY ANN" "THOMAS", or "GEORGE" "HERBERT" "WALKER" "BUSH" becomes "GEORGE" "HERBERT WALKER" "BUSH"); or
- any combination of the above.

22. I am a member of the Association for Computing Machinery, and I have reviewed relevant portions of the ACM's February 2006 study *Statewide Databases of Registered Voters*, a copy of which is attached as Exhibit H. This study recognizes both that "[m]ost errors in individual database records occur during data entry," and that "[w]hile quality control systems and appropriate supervision of data entry may reduce data entry errors, some errors will inevitably occur. . . . Changes that are primarily entered in other state databases – such as changes in marital status and court approved name changes – also compound the challenge to accuracy." Exh. H at 21.

23. In my extensive experience working with databases containing similar kinds of personal information, the errors described in paragraphs 20 and 21 can be quite

HILLIS CLARK MARTIN & PETERSON, P.S.
500 Galland Building, 1221 Second Ave
Seattle WA 98101-2925
206.623.1745; fax 206.623.7789

common. One reliable study found that the names of 23-37% of the patients in several medical databases were misspelled in at least one database record; a copy of this study is attached as Exhibit I.

24. If any one or more of the errors described in paragraphs 20 and 21 were to occur in Washington in the registration record itself and/or in the database with which the record will be matched, the name as entered from the individual's registration record will not exactly match the name as recorded in the database with which the State will compare the individual's registration information.

25. Data entry operators commonly commit errors when they input names, but they also commit many of the same types of errors when they input numbers. Such errors are specifically acknowledged to occur with respect to Social Security Numbers. The leading expert on record matching for the U.S. Bureau of the Census estimates that in one large California employment database, given these types of errors "[o]ver a period of twenty years, the records [associated] with each individual can expect to contain *at least two errors* where the [Social Security Number] has been mis-keyed or transcribed improperly" (emphasis added). A copy of this publication is attached as Exhibit J.

26. **Data Maintenance, Storage, Transfer, and Transformation.** Once a record is created for an individual applicant, the State must maintain, store, transfer and, often, transform the data contained in that record. Federal and State officials must perform similar tasks with respect to data contained in the Social Security Administration and Department of Licensing databases. These processes are also prone to error, for example, when computer viruses cause file corruption; when the data input locally, in Washington's 39 different county election management systems, is transferred to the State; and when database fields are added, modified or deleted and, accordingly, data is split, changed, or consolidated. In my experience, such transfers can lead to unintended changes in the underlying data. For example, in the process of transferring a file from one database to another, one of my clients found that all spaces in all address fields had been unintentionally deleted.

HILLIS CLARK MARTIN & PETERSON, P.S.

500 Galland Building, 1221 Second Ave
Seattle WA 98101-2925
206.623.1745; fax 206.623.7789

27. The ACM's study, *Statewide Databases of Registered Voters*, also recognizes that glitches can create problems in large databases. As the study states:

> Databases also can be inaccurate or unreliable because of computer viruses, programming errors, and system failures. For example, in 2003 the Maryland Motor Vehicle Administration (MVA) offices were attacked by a computer worm. The worm shut down the MVA's computers and telecommunication systems, cutting them off from all forms of remote communication and disrupting operations in all 23 MVA offices located throughout the state. A second event occurred on January 20, 2004, when the MVA could not process work on the mainframe computer for about an hour after opening. The problem was characterized as a computer glitch.

Exh. H at 24.

28. There is no single standard industry algorithm or process for maintaining, storing, or transforming data; different entities use different processes for these purposes. As noted on page 14 of the *"Social Security Verification" System Specification*, for example, there will be "many different types of computers on the [AAMVA] network, each possibly having a different data-encoding scheme." Different entities using different conventions, or transferring data using different encoding systems may, because of incompatibilities, cause modifications in the data they maintain that will lead to unmatched information.

29. If any of these modifications were to occur in Washington, affecting the registration record itself and/or the database the record will be matched with, the information as entered from the individual's registration record will not exactly match the information as recorded in the database with which the State will compare her identifying information.

30. <u>System Errors.</u> Online computer systems intermittently experience system errors or other "down time." The Social Security Administration is not immune to these errors; the *"Social Security Verification" System Specification* describes "program problems, network interface errors, database errors, program aborts, [and] the more common system error[,] when the SSA file is off-line." Exh. B at 17.

HILLIS CLARK MARTIN & PETERSON, P.S.

500 Galland Building, 1221 Second Ave
Seattle WA 98101-2925
206.623.1745; fax 206.623.7789

31. I have no information about the practice in Washington state regarding such errors. However, if such system errors occur when a Washington registration record is submitted, at least during the error period, the information on that record will not be able to be matched with information in the offline database.

32. **Natural Data Inconsistency.** In addition to the errors described above, the process of matching information in different records itself produces false negatives because of superficial discrepancies between those records that do not reflect inaccurate information. For example, names are not truly standardized, nor are they fixed. People adopt nicknames, use shortened names, pick up or drop middle names, take their spouse's names, and/or change the spelling of their transliterated names – and they do so even in formal government documents. In addition, different applicants or different data entry operators (and even the same people on different occasions) may transliterate non-English characters in different ways. Thus, two records for the same person may show different names, like a maiden name or married name. Similarly, data entry operators often use default assumptions to fill in missing information (*e.g.*, choosing the first of the month when no day of the month is given).

33. Common examples of natural data inconsistencies that may cause false negatives include:

- nicknames (*e.g.*, "SAMUEL" versus "SAM");
- maiden names (*e.g.*, "MARY JONES" versus "MARY SMITH");
- husband's names (*e.g.*, "MRS. JOHN SMITH" versus "MRS. MARY SMITH");
- punctuation (*e.g.*, "O'BRIEN" versus "O BRIEN" or "OBRIEN")
- compound last names (*e.g.*, "HILLARY" "RODHAM CLINTON" versus "HILLARY" "RODHAM" "CLINTON");
- first or middle initials (*e.g.*, "F. SCOTT FITZGERALD" versus "FRANCIS S. FITZGERALD");
- name change due to religious conversion (*e.g.*, "MUHAMMAD ALI" versus "CASSIUS CLAY"); or

HILLIS CLARK MARTIN & PETERSON, P.S.
500 Galland Building, 1221 Second Ave
Seattle WA 98101-2925
206.623.1745; fax 206.623.7789

- any combination of the above.

34. Similar data inconsistencies arise when confronting names common within certain ethnic communities:

- immigrants adopting "Americanized" names, for all purposes or just some purposes (*e.g.*, "GRACE KIM" versus "HYUN KIM");
- name change due to different status in the community (*e.g.*, in Burmese, "MAUNG TIN" (for younger men) versus "U TIN" (for married men));
- mistaking a title for a first name (*e.g.*, "MAUNG TIN" versus "TIN");
- transliterated names or diacriticals (*e.g.*, "MUHAMMAD" with "MOHAMMED," or "JÜRGEN" with "JUERGEN" or "JURGEN," or "HWANG" with "WHANG" or "WANG"); or
- any combination of the above.

35. In my extensive experience working with databases containing similar kinds of personal information, the discrepancies described in paragraphs 33 and 34 can be quite common. If any of the natural discrepancies described in paragraphs 33 and 34 were to occur in Washington, creating immaterial differences between the registration record itself and the database the record will be matched with, the information as entered from the individual's registration record will not exactly match the information as recorded on the database with which the State will compare her identifying information.

36. My wife's name provides a good example of how such trivial differences can cause problems in record-matching. My wife usually represents herself as "Sarah C. Borthwick," and signs personal checks that way. But she is registered to vote in New York as "Sarah E. Caguiat Borthwick." Her New York driver's license shows her name as "Caguiat-Borthwick, S". And she appears in Social Security Administration records as "Sarah E. Caguiat." If she attempted to register to vote in Washington as "Sarah Borthwick," the information in her application would likely not match either information in either the driver's license or Social Security databases.

HILLIS CLARK MARTIN & PETERSON, P.S.
500 Galland Building, 1221 Second Ave
Seattle WA 98101-2925
206.623.1745; fax 206.623.7789

Errors Common in Particular Communities

37. Certain errors contributing to difficulties in record-matching are more prevalent among particular racial and ethnic communities. For example, transposition of the "first" name and "last" name is more common with regard to individuals of Chinese descent, many of whom present their family name first and their given name second, contrary to the usual American practice. A data entry operator might not know which name in "Lu Bao" is the first name and which is the second, and enter it based on any variety of conventions, such as assuming that the first name listed is the given name. If Mr. Lu's name is transposed in one record, that name will not match exactly to the other record. In my experience, individuals of Chinese descent also frequently adopt names considered to be common "Western names," but use these "Western names" inconsistently in official records. The following example illustrates both phenomena: a Chinese woman named "Wang Fei" might inconsistently put her first name before her last name (*i.e.*, "Fei Wang"); use a Western form of her first name (*e.g.*, "Faye Wang") or a Western name not derived from her first name (*e.g.*, "Grace Wang"); and/or use a Western form for both her first and last names (*e.g.*, "Faye Wong").

38. In African-American communities, names derived through modification of more traditional spellings are more common than in other racial or ethnic communities. These names are more likely to be misspelled in data entry. For example, one study reports that "Jazmine," "Jasmin," and "Jazmin" are all girls' names much more common among African-Americans. A copy of this study is attached as Exhibit K. These names may all be misspelled as "Jasmine" in data entry, thus creating errors when an exact character-by-character match protocol is applied. Moreover, names that are unique to a particular individual are also more common in African-American communities. The same study cited above, for example, found that African Americans in California are six times more likely to have a unique name than are Caucasians. These names may be unfamiliar to data entry personnel (of any race or ethnicity), and are more likely to be misspelled in a database.

HILLIS CLARK MARTIN & PETERSON, P.S.

500 Galland Building, 1221 Second Ave
Seattle WA 98101-2925
206.623.1745; fax 206.623.7789

39. In Hispanic or Latino communities, it is common to use either maternal or paternal last names, or both. These names are often supplied inconsistently by the individual or entered inconsistently by the data entry operator such that the "middle name" and "last name" fields in the resulting record are inverted, separated, or combined. For example, "José Luis Rodriguez Zapatero" might have "Zapatero," "Rodriguez," or "Rodriguez Zapatero" entered as his last name.

40. In communities that do not use the Roman alphabet in their primary language, such as East Asian, Middle Eastern, Hellenic, and Slavic communities – or communities using diacritical marks not found in English, such as the umlaut or tilde – inconsistent transliterations are common. Arabic names like "Mohammed," for example, are transcribed differently depending on the country of origin. Three variants include "Muhammad," "Mohamed," and "Mahomet."

41. The transposition of the date and month of birth is more commonly found with regard to recent immigrants, who may be accustomed to presenting dates in a day-month-year convention, which is commonly used in Europe, Africa, the Middle East, and Asia. Thus, someone whose date of birth is May 6, 1980 might input her name as "6/5/1980," and since that is a valid date under the American month-day-year convention, her record will reflect that her birth date is June 5, 1980.

42. Mismatched surnames due to a maiden name or married name are more common, of course, with regard to women. Thus, to use my wife as an example again, whether she registers to vote using a compound last name without a hyphen, a compound last name with a hyphen, or my last name, the information entered from her registration record will not exactly match the information in the Social Security Administration's database, where she appears under her maiden name.

The Impact of Errors and Non-Standardized Data on Record Matching

43. Attempts to match records using exact, character-by-character matching – referred to in the industry as "deterministic" matching – are highly sensitive to all of the errors

HILLIS CLARK MARTIN & PETERSON, P.S.
500 Galland Building, 1221 Second Ave
Seattle WA 98101-2925
206.623.1745; fax 206.623.7789

and discrepancies described above. The failure to match information because of such errors and discrepancies would result in false negatives – *i.e.*, the failure to match database entries that in fact belong to the same individual.

44. Data from several reliable studies show that, in similar circumstances, false negative rates generated by deterministic matching protocols can reasonably be expected in the range of 20-30%. For example, in one reliable study, the U.S. Bureau of the Census suggested that using a deterministic match on census data would have resulted in a false negative rate of about 25%; a copy of this study is attached as Exhibit L. Another reliable health care study found a false negative rate of about 22% using a deterministic protocol. Exh. I at 503-04. And yet another reliable study found that a deterministic protocol missed 17-30% of records belonging to the same individuals; a copy is attached as Exhibit M.

45. As noted above, I have reviewed the *Social Security Verification System Specification* prepared by the American Association of Motor Vehicle Administrators in August 2004, and, in particular, the HAVV transaction described on pages 26 and 27 of that document. Exh. B at 26-27. As described in that document, the HAVV transaction uses a deterministic match protocol in which a system will attempt to match the last name, first name, month of birth, year of birth, and last four digits of the Social Security number of a target record to the same elements of records in the Social Security Administration database. A successful match will be reported only when each character of each such field in the target record matches precisely each character of each corresponding field in the Social Security Administration database. Pursuant to the same document, I understand that an unsuccessful match will be coded as a "system error," "invalid input data," or "no match found"; no more specific information will be returned to the state indicating why a match could not be found, or more precisely locating the source of the error.

46. The HAVV protocol is not designed to account for, and will not readily account for, the errors described above. Moreover, the requirement that *multiple* fields exactly match compounds the error rate expected for an exact match on any individual field.

HILLIS CLARK MARTIN & PETERSON, P.S.

500 Galland Building, 1221 Second Ave
Seattle WA 98101-2925
206.623.1745; fax 206.623.7789

47. Page 10 of the presentation noted above of Mr. Monaghan, the Social Security Administration's Director of Information Exchange, states that no match was found in 28.5% of 143,000 queries submitted in the period before his presentation. Exh. D at 10.

48. Assuming that the Social Security Administration used the deterministic HAVV transaction to seek matches for voter registration records, the reported 28.5% "no match" rate described in paragraph 47 is consistent with the rate of false negatives found in other published accounts of deterministic matching. That is, it would be consistent with these accounts to find that 28.5% of the queries submitted to the Social Security Administration failed to match information in the Social Security database, but actually represent individuals accounted for in the Social Security database.

49. Moreover, in my opinion, some voters will not be provided an effective opportunity to resolve a "false negative." For example, although I understand that Washington election officials are required to attempt to correspond with unmatched registrants, data entry errors impacting name and address will prevent some correspondence from reaching its intended target.

50. I have reviewed the May 15, 2003 appraisal of Virchow Krause & Company, a prominent Midwest accounting and consulting firm retained by the Wisconsin State Elections Board to evaluate project proposals for Wisconsin's statewide voter registration database. A copy of the relevant portion of this appraisal is attached as Exhibit N. I agree with the appraisal's conclusions regarding the difficulty and likely effect of matching in this context:

> Name matching and validation issues are very complex (e.g., matching Margie L. Smith with Margaret Smith), and are made even more complex when aliases and name changes are considered. . . . Even a 1% error rate on an interface validating names, driver license numbers, etc. could generate tens of thousands of bad matches in an error log, well beyond any ability for the [state, county, or local] users to manually verify the errors. . . . [¶] All vendors suggested that incomplete or unmatched records be ignored, because the time to resolve, cost to resolve, and potential for error and disenfranchisement was too high.

Exh. N at 20.

HILLIS CLARK MARTIN & PETERSON, P.S.

500 Galland Building, 1221 Second Ave
Seattle WA 98101-2925
206.623.1745; fax 206.623.7789

In sum, the matching systems that I understand Washington is using, described in paragraphs 15 and 45, are prone to many errors – especially false negatives. In my opinion, such systems would generate failed matches for individuals who are, in fact, legitimately represented in the target database. I would not use those matching systems – or any similar system. When comparing two data sources of significant size – as Washington is doing here – records representing the same individual would fail to match even if the Secretary of State used protocols representing the best available technology. If matching is a prerequisite to registration, the use of any match process will result in eligible voters being denied the right to vote.

I declare under penalty of perjury under the laws of the United States of America that the foregoing is true and correct, and that this Declaration was executed on May 24, 2006 in New York, New York.

ANDREW BORTHWICK

#350293 99732-002 72%t01!.doc 5/24/2006

HILLIS CLARK MARTIN & PETERSON, P.S.

500 Galland Building, 1221 Second Ave
Seattle WA 98101-2925
206.623.1745; fax 206.623.7789

CERTIFICATE OF SERVICE

I hereby certify that on May 24, 2006, I ~~electronically~~ filed this ***Declaration of Andrew Borthwick*** with the Clerk of the Court ~~using the CM/ECF system~~ which will send notification of such filing to the following:

N/A

and I hereby certify that I have sent for service via hand delivered by legal messenger to be served on May 25, 2006 this document to the following non CM/ECF participants:

Sam Reed, Secretary of State, State of Washington
Legislative Building
Olympia, WA 98504-0220

Rob McKenna, Attorney General for the State of Washington
Office of the Attorney General
1125 Washington Street SE
Olympia, WA 98504-0100

I declare under penalty of perjury under the laws of the United States of America that the foregoing is true and correct.

DATED this 24th day of May, 2006 at Seattle, Washington.

/S/ Sarah A. Dunne // Sarah A. Dunne
Sarah A. Dunne, WSBA #34869

HILLIS CLARK MARTIN & PETERSON, P.S.
500 Galland Building, 1221 Second Ave
Seattle WA 98101-2925
206.623.1745; fax 206.623.7789

Exhibit A

ANDREW BORTHWICK
48 Wall Street, 11th Floor
New York, NY 10005
Phone: (212) 918.4412
Email: andrew@choicemaker.com

EXPERIENCE

CHOICEMAKER TECHNOLOGIES, INC., New York, NY — July 1998-Present

CEO and Co-Founder

- Founded the company to commercialize an innovative solution to the problem of record duplication based on my Ph.D. research.
- Built the company from nothing to $1.7M in annual revenues.
- Coordinated the successful deployment of systems with the New York City Department of Health, NY, IA, KS, MO, NE, KS, SC, and WY State Education Departments, Regulatory DataCorp, Phoenix-ESI, and other clients.
- Hands-on work in marketing, including helping develop marketing collaterals and the creation of design and content for the website.
- Actively involved in sales, including technical sales support and contract negotiation.
- Give invited talks and numerous publications and presentations on record matching to help promote the firm.
- Manage the firm's finances. Carefully monitored cash flow to enable firm to grow on minimal equity investments. Secured equity, debt, and government grant financing for the firm.
- Lead all HR activities, including hiring, evaluation, and dismissal.
- Principal investigator of three National Science Foundation Small Business Innovation Research grants totaling $1M for research into artificial intelligence approaches to approximate record matching.
- Awarded US Patent 6,523,019 for machine learning approach to record matching. Co-inventor of two other pending patents.
- Personally coded ChoiceMaker version 1.0.

MORGAN STANLEY, New York, NY — 1993-2002

Systems Consultant

- Working only one day per week, was critical designer and maintainer of the Information Services Allocation Model (ISAM). ISAM is a highly sophisticated system which solves matrix algebra equations describing the circular movement of money within IT in order to equitably allocate over $1 billion in annual IT costs to the rest of the firm.
- Designed most of the major and minor upgrades for ISAM, which were implemented by a team of three full-time programmers. The system grew greatly in functionality and importance over nine years.
- Duties included making presentations to explain the functionality of the existing system, clarifying user requirements, designing enhancements, answering user questions, fixing bugs, and Y2K.
- One of a small number of consultants put on a *must-retain* list during a switchover of consulting agencies.

IBM WATSON LABORATORY, Yorktown Heights, NY — Summer 1997

Summer Intern

- Researched maximum entropy language modeling for a voice-operated air travel reservation system.

MORGAN STANLEY, New York, NY — 1988-1993

Programmer/Technical Designer/Business Analyst

- Designed and managed the project which built ISAM.
- Supervised a team of four while working closely with the users in IT Finance to take the project from a few pages of notes, diagrams, and equations to a finished product.
- Personally coded the mathematical heart of the system.

EDUCATION

New York University, New York, NY — September 1999

- Ph.D., Computer Science
- Invented and constructed a system to detect proper names ("named entities") in newspaper text. Built the first system to combine the output of multiple hand-coded information extraction systems within a maximum entropy framework. System placed fourth out of twelve in a DOD evaluation after only four person-months of effort. Rapidly ported the system to Japanese and performed well in a Japanese named entity evaluation, where it was the only system written by a non-speaker of Japanese.
- Thesis title: "A Maximum Entropy Approach to Named Entity Recognition"

- Specialized in Artificial Intelligence and Natural Language Processing

Oberlin College, Oberlin, OH — May 1988
- Bachelor of Arts, History
- Phi Beta Kappa
- Comfort Starr Prize for Excellence in History

PUBLICATIONS

- Andrew Borthwick. *The Design and Testing of a Record Matching System*. Slides and abstract. 17th Information Quality Conference. Houston, Texas. September 21, 2005.
- Andrew Borthwick and Maggie Soffer. *Business Requirements of a Record Matching System*. Peer reviewed paper. Massachusetts Institute of Technology's Ninth International Conference on Information Quality (MIT ICIQ), Cambridge, MA. September 7, 2004.
- Martin Buechi, Andrew Borthwick, Adam Winkel, and Arthur Goldberg. *ClueMaker: A Language for Approximate Record Matching*. Peer reviewed paper. Massachusetts Institute of Technology's Eighth International Conference on Information Quality (MIT ICIQ), Cambridge, MA. August 27, 2003.
- Andrew Borthwick, Vikki Papadouka, and Deborah Walker. *Advanced Name-Matching Techniques for Immunization Registries*. "Immunization Registry Conference", Slides and abstract, July 2001.
- Andrew Borthwick and Deborah Walker. *Applications of Record Matching Techniques for a Lead-Immunization Registry Integration Project*. "35th National Immunization Conference", Slides and abstract, Atlanta, Georgia, May 2001.
- Andrew Borthwick, Vikki Papadouka, and Deborah Walker. *Principles and Results of the New York Citywide Immunization Registry's MEDD Deduplication Project*. "34th National Immunization Conference", Slides and abstract, Washington, DC, July 2000.
- Andrew Borthwick, Vikki Papadouka, and Deborah Walker. *The MEDD Deduplication Project*. "2000 Immunization Registry Conference", Slides and abstract, Newport, Rhode Island, March 2000.
- Andrew Borthwick, Vikki Papadouka, and Deborah Walker. *New Techniques for Registry Deduplication in the MEDD Deduplication Project*. "2000 Immunization Registry Conference", Slides and abstract, Newport, Rhode Island, March 2000.
- Andrew Borthwick. *A Probabilistic Record Linkage Model Derived from Training Data*. U.S. Patent #6,523,019. Filed Oct. 28, 1999. Awarded February 18, 2003.
- Andrew Borthwick. *A Maximum Entropy Approach to Named Entity Recognition*. PhD thesis, New York University, New York, New York, September 1999.
- Andrew Borthwick. *A Japanese Named Entity Recognizer Constructed by a Non-Speaker of Japanese*. "Proceedings of the IREX Workshop", Tokyo, Japan, August 1999.
- Andrew Borthwick, Vikki Papadouka, Deborah Walker, Amy Metroka, Charles Troob, and Stephen Friedman. *The MEDD Deduplication Project*. "Slides and abstract, 33rd National Immunization Conference", Dallas, Texas, June 1999.
- Andrew Bothwick, John Sterling, Eugene Agichtein, and Ralph Grishman. *Exploiting Diverse Knowledge Sources via Maximum Entropy in Named Entity Recognition*. "Proceedings of the Sixth Workshop on Very Large Corpora", August 1998.
- Andrew Borthwick, John Sterling, Eugene Agichtein, and Ralph Grishman. *NYU: Description of the MENE Named Entity System as used in MUC-7*. "Proceedings of the Seventh Message Understanding Conference (MUC-7)", Fairfax, Virginia, April 1998.

TECHNICAL EXPERTISE

Operating Systems
- Windows, Unix (Linux and Sun), IBM Mainframe

Technologies
- Data quality, computational linguistics, graph algorithms, software architecture

Programming Languages
- ClueMaker, C/C++, Perl, Adabas/Natural, Java, APL

Software Packages
- ChoiceMaker 2, Eclipse, MS Project, Visio, Quickbooks, MS Office