

ARIZONA SUPREME COURT

DAVID LAKE,

Plaintiff/Appellant/  
Petitioner,

v.

CITY OF PHOENIX, a political  
subdivision of the State of Arizona;  
FRANK FAIRBANKS, in his official  
capacity; MARIO PANIAGUA, in his  
official capacity; JACK HARRIS, in his  
official capacity,

Defendants/Appellees.

Supreme Court  
CV-09-0036-PR

Court of Appeals  
Division One  
No. 1 CA-CV -7-0415

Superior Court  
Maricopa County  
No. LC2006-00835-001 DT

**DECLARATION OF STEPHEN K. DOIG IN SUPPORT OF BRIEF OF  
*AMICI CURIAE* FIRST AMENDMENT COALITION OF ARIZONA, INC.,  
SOCIETY OF PROFESSIONAL JOURNALISTS, AND ARIZONA  
NEWSPAPERS ASSOCIATION**

I, Stephen K. Doig, declare as follows:

1. I am a professor at the Walter Cronkite School of Journalism and Mass Communication at Arizona State University, where I hold the Knight Chair in Journalism specializing in computer-assisted reporting.

2. Prior to joining the faculty of the Cronkite School, I spent more than 19 years as an investigative reporter at *The Miami Herald*. During my years at the *Herald*, I became a nationally known expert on the use of database and statistical software to analyze computer-readable public record datasets. Significant investigative projects on which I worked at the *Herald* have been recognized with a variety of major awards, including the 1993 Pulitzer Prize for Public Service, the Investigative Reporters & Editors' Award, Harvard University's Goldsmith Prize for Investigative Reporting, the American Bar Association's Silver Gavel award, and the Inter-American Press Association award.

3. Since joining the Cronkite School, I have remained active in the journalism profession. I do frequent training sessions for reporters who wish to learn how to analyze public record datasets. I have trained hundreds of reporters around the United States. I also have conducted computer-assisted reporting workshops in Spain, Brazil, Indonesia, Norway, the Netherlands, Belgium, England, Canada and Mexico.

4. I served four years as an elected member of the board of directors of Investigative Reporters & Editors, a professional organization of more than 4,000 members. I also organized and judge the annual Phil Meyer Award for Precision Journalism, which recognizes the best reporting done using social science methods.

5. The metadata of a dataset contains detailed information that is essential for reading and analyzing the data properly. Key elements of metadata include:

- The file layout, which explains how to parse each line of the data into the proper variables.
- The data dictionary, which explains the codes and abbreviations necessary to understanding what the variables mean.
- The relational architecture, which show how different tables of a relational database can be joined on common variables

6. Metadata also can include other information about a computer file, including revision history, contact information for those responsible for managing the data, software used to create it, map projections used in geographic files, etc.

7. Here is a hypothetical example of two records from a very simple computer dataset:

194804211722388500410

195008030691728528121

8. The metadata shown below for this dataset in paragraph 7 above, combines the file layout and the data dictionary:

<u>Column</u>	<u>Variable</u>
1-4	Birth Year
5-6	Birth Month
7-8	Birth Day
9	Gender (0=Female, 1=Male)
10-11	Height in inches
12-14	Weight in pounds
15-19	ZIP Code
20	Race (1=White, 2= Black, 3=Other)
21	Hispanic Origin (1=Yes, 0=No)

9. Applying the metadata in paragraph 8 to the two datasets in paragraph 7, we learn that the first set of numbers describes a White male who was born on April 21, 1948, is 6 feet tall, 238 pounds and lives the downtown Phoenix zip code of 85004. The second set of numbers in paragraph 7 describes a black Hispanic female who was born on August 3, 1950, is 5 feet, 9 inches tall, 172 pounds and lives in the Tempe zip code of 85281.

10. The records in the dataset cannot be interpreted or used without the metadata. Refusal to supply the metadata would be tantamount to denying the

public records request. The purpose of the Public Records Law is to allow citizens to monitor the workings of government. Refusing to include metadata when producing a dataset pursuant to a public records request is equivalent to releasing data in a password-protected file without revealing the password. Attached as an appendix is the metadata for a much more complicated relational database

11. Metadata itself poses no problems of confidentiality or invasion of privacy. The metadata indeed may list one or more variables that themselves contain information that is confidential, such as the identity of confidential informants in a police database. But the variable name (perhaps in this example “CI” or “CONF\_INF”) reveals nothing damaging itself, and it is a simple matter for the record custodian to export the database so that the problematic variable is excluded. Refusing to produce the metadata for a dataset that contains some confidential variables is similar to refusing to produce any police arrest report that contains the name of a confidential informant, even though those names could readily be redacted with a black marker.

12. Metadata routinely is used by journalists to analyze government data in order to produce reports of vital public interest. A few of the hundreds of examples include:

- Analysis of voting records by the *Miami Herald* in a 1998 Miami mayoral election that proved pervasive election fraud and caused the election to be overturned.
- Analysis of stock option grants by the *Wall Street Journal* that showed dozens of public companies were backdating options for favored insiders to give them inflated profits.
- Analysis of Florida teacher discipline records by the *Sarasota Herald Tribune* showing how predatory teachers were protected by a system that allowed them to move from school to school.
- Analysis by the *New York Times* of fatal accidents at railway crossings across the nation documenting corporate cover-up of responsibility.
- Analysis of mortgage lending practices in Atlanta by the *Atlanta Journal-Constitution* that demonstrated how banks were rejecting loan applications from black homebuyers at much higher rates than whites with similar financial resources.

13. The project for which I won the Pulitzer Prize featured an analysis of the damage patterns to more than 80,000 homes in south Florida after Hurricane Andrew in 1992. I used computer-readable datasets that included the county's property tax roll of more than 600,000 records; more than a million building

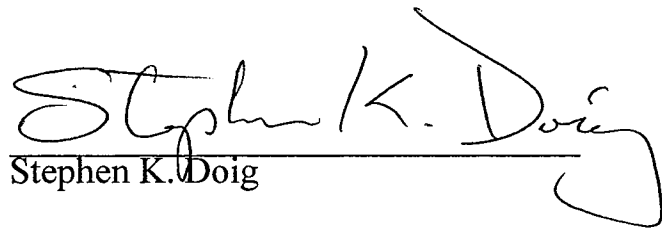
inspection records; and tens of millions of dollars in campaign finance records of showing the influence of housing developers on the county commission. The analysis demonstrated that weakening of county building codes, sought by developers and approved by commissioners, greatly magnified the financial and human disaster of the storm. This project could not have been accomplished without access to the metadata of the various complex datasets I used. The metadata for each of these datasets encompassed many pages detailing how to import the records and explaining the meaning of various codes used by the property appraisers and building inspection officials.

14. Withheld or inaccurate metadata can lead to errors in interpretation. An example of this problem occurred when I did an analysis of how south Florida judges were sentencing people convicted of driving under the influence. The metadata told me there was a variable that showed how much jail time each defendant received and another that showed how large a fine each received. My analysis of the data showed for each judge the percentage of cases that were given jail and the percentage given fines. Every judge also had 1-2% of his or her cases showing no jail and no fine, a finding that was included in the graphic. When the project was published in the *Miami Herald*, the judges complained that the graphic indicated they were breaking the law – all DUI cases must receive punishment. As I

learned later in talking with the court clerks' office, zero-jail-zero-fine defendants actually were indigents on their first offence who were given community service as punishment. Because the metadata was flawed, this fact was not included in the metadata and therefore led to an error that had to be corrected

I declare under penalty of perjury that the foregoing is true and correct.

Executed on July 14, 2009 in Phoenix, Arizona.

  
Stephen K. Doig

29047-0001/LEGAL16474980.1

# APPENDIX

## Campaign Finance Database Table Specification

At the end of the 1998 cycle, data from all open committees was imported from the existing database to a new database that worked in conjunction with our new Campaign Finance Software. This document describes the tables in the new database.

Revised November 11, 2001

- ORGS table, STATUS and TERM\_DATE added (previously these were implied by filings)
- Table EFILINGS table, as received electronic filings audit log
- Table CECPERIODS, potential reporting dates for candidate committees
- Table ELECTIONDATES, holds the dates of the primary and general for the CYCLE
- Table LIMITS, defines the limits by which a candidate must operate within a CYCLE
- Table ORG\_HIST, as changed ORG information audit log

Revised July 17, 2000

- ORGS table, CANDINCUMBENT renamed to CANDINCUMBANT
- ORGS table, CANDOFFICE and CANDCOUNTY renamed to CANDOFFICEID and CANDCOUNTYID respectively.
- Table Namelist had the PHONE and ZIP swapped in order. Correct order is ZIP, then PHONE.

### Table Descriptions:

Each table's first record is a HEADER record.

The major primary key is FILERID which is the unique ID number assigned to each committee by the Secretary of State.

## ORGS

Main table defining the committee. This table only holds two records, the HEADER record and the record of the committee filing.

/\* Table: ORGS \*/

```
CREATE TABLE ORGS
  FILERID                INTEGER                /* UNIQUE IDENTIFIER*/
  COMMITTEETYPEID       INTEGER                /* POINTER TO COMMITTEE TYPE*/
  NAME                   VARCHAR(90)
  ADDRESS1               CHAR(40)
  ADDRESS2               CHAR(40)
  CITY                   CHAR(20)
  STATE                  CHAR(2)
  ZIP                    CHAR(10)
  MAILADDRESS1           CHAR(40)
  MAILADDRESS2           CHAR(40)
  MAILCITY                CHAR(20)
  MAILSTATE              CHAR(2)
  MAILZIP                CHAR(10)
  PHONE                  CHAR(14)
  FAXPHONE               CHAR(14)
  EMAIL                  CHAR(20)
  CHNAME                 CHAR(50)            /* CHAIRMAN'S NAME*/
  CHADDRESS1             CHAR(40)
  CHADDRESS2             CHAR(40)
  CHCITY                  CHAR(20)
  CHSTATE                CHAR(2)
  CHZIP                  CHAR(10)
  CHPHONE                CHAR(14)
  CHFAXPHONE             CHAR(14)
  CHOCCUPATION           CHAR(50)
  CHEMPLOYER             CHAR(50)
  TRNAME                 CHAR(50)            /* TREASURER'S NAME*/
  TRADDRESS1             CHAR(40)
  TRADDRESS2             CHAR(40)
```

TRCITY	CHAR(20)	
TRSTATE	CHAR(2)	
TRZIP	CHAR(10)	
TRPHONE	CHAR(14)	
TRFAXPHONE	CHAR(14)	
TROCCUPATION	CHAR(50)	
TREMPLOYER	CHAR(50)	
CANDNAME	CHAR(50)	/* CANDIDATE'S NAME*/
CANDADDRESS1	CHAR(40)	
CANDADDRESS2	CHAR(40)	
CANDCITY	CHAR(20)	
CANDSTATE	CHAR(2)	
CANDZIP	CHAR(10)	
CANDPARTY	CHAR(30)	
CANDOFFICEID	INTEGER	/* REFERENCE FOR EXT. TABLE "OFFICE CODE" (OFFCODE.TXT) */
CANDCOUNTYID	INTEGER	/* REFERENCE FOR EXT. TABLE "COUNTY CODE" (COUNTIES.TXT) */
CANDINCUMBANT	CHAR(1)	/* (T)RUE / (F)ALSE*/
CANDPHONE	CHAR(14)	
CANDFAXPHONE	CHAR(14)	
CANDEMAIL	CHAR(14)	
CANDOFFICENAME	CHAR(20)	/* DESC. FROM EXT. TABLE "OFFICE CODE" (OFFCODE.TXT) */
CANDYEAR	SMALLINT	/* CANDIDATE RUNNING YEAR*/
DESIGNEENAME	CHAR(50)	/* DESIGNEE OF AUTHORITY NAME*/
DESIGNEEADDRESS1	CHAR(40)	
DESIGNEEADDRESS2	CHAR(40)	
DESIGNEECITY	CHAR(20)	
DESIGNEESTATE	CHAR(2)	
DESIGNEEZIP	CHAR(10)	
DESIGNEEPHONE	CHAR(14)	
DESIGNEEFAXPHONE	CHAR(14)	
DESIGNEEMAIL	CHAR(50)	
SPONSOR	VARCHAR(50)	/* SPONSOR'S NAME*/
SPONSORADDRESS1	CHAR(40)	
SPONSOR CITY	CHAR(20)	
SPONSOR STATE	CHAR(2)	
SPONSOR ZIP	CHAR(10)	
SPONSORRELATIONSHIP	CHAR(40)	
REG_DATE	DATE	/* REGISTRATION DATE*/
AMEND_DATE	DATE	/* AMMENDMENT DATE*/
ALLOWEXCESS	CHAR(1)	
EXCESS_DATE	DATE	
OUTOFSTATE	CHAR(1)	/* OUT OF STATE MARKER*/
CECAPPLYDATE	DATE	/* DATE COMMITTEE APPLIES FOR PARTICIPATION*/
CECDATE	DATE	/* DATE COMMITTEE PARTICIPATES*/
NOACTIVITYDATE	DATE	
LASTREPORTDATE	DATE	
TRIGGERLEVELPRIMARY		
TRIGGERLEVELGENERAL		
GENREPORTFLAG		
GENREPORTDATE		
MODIFIED_DATE	DATE	/* LAST MODIFIED DATE*/
MODIFIED_USER	CHAR(20)	/* LAST MODIFIED BY USER*/
DELETED	CHAR(1)	/* (T)RUE or (F)ALSE*/
FINANCIALINSTITUTION1	/* CHAR(20) */	
FINANCIALINSTITUTION2	/* CHAR(20) */	
FINANCIALINSTITUTION3	/* CHAR(20) */	
REALLOCATIONFLAG	CHAR(1)	/* (Y)ES / (N)O*/
STATUS	CHAR(1)	/* (A)ctive / (I)nactive */
TERM_DATE	DATE	/* Date of Termination */

## **ORGS\_HIST**

Audit history for tracking the committee. This table is the same as the ORGS table, with a CHANGESEQ to represent the order of change.

## COMTYPE

Type of committee

COMTYPE(COMMITTEETYPEID)

→ ORGS(COMMITTEETYPEID)

/\* Table: COMTYPE \*/

CREATE TABLE COMTYPE

COMMITTEETYPEID	INTEGER	UNIQUE IDENTIFIER
COMMITTEETYPEDESCRIPTION	CHAR(100)	DESCRIPTION OF COMMITTEE PURPOSE
COMMITTEECANDIDATE	CHAR(1)	T or F for candidate committee

## FILINGS

Filings made by each committee, defined by cycle and period of filing

FILINGS(FILERID)

→ ORGS(FILERID)

/\* Table: FILINGS \*/

CREATE TABLE FILINGS

FILERID	INTEGER	POINTER TO ORGS TABLE
CYCLE	INTEGER	CAMPAIGN CYCLE
PERIODNUMBER	INTEGER	REPORTING PERIOD
DATE_FILED	DATE	DATE FILED
RELEASE	CHAR(1)	RELEASE TO WEB
DATE_AMENDED	DATE	DATE AMENDED
AMOUNTPAID	DOUBLE	LATE FILING FEE
ELECTRONIC	CHAR(1)	FILED ELECTRONICALLY
NOACTIVITY	CHAR(1)	DEFAULT "F"

## EFILINGS

Electronic Filings made by each committee as received by date

EFILINGS(FILERID)

→ ORGS(FILERID)

/\* Table: FILINGS \*/

CREATE TABLE EFILINGS

FILERID	INTEGER	POINTER TO ORGS TABLE
FILINGDATE	DATE	DATE FILED
FILENAME	CHAR(255)	FILE NAME
MODIFIEDDATE	DATE	DATABASE MAINTAINED RECORD MODIFIED
RECEIPT	CHAR(10)	RECEIPT TO USER

## PERIODS

Period lookup (cycle, periodnumber)

PERIODS(CYCLE,PERIODNUMBER)

→ FILINGS(CYCLE,PERIODNUMBER)

/\* Table: PERIODS \*/

CREATE TABLE PERIODS (

CYCLE	INTEGER	CAMPAIGN CYCLE (yyyy)
PERIODNUMBER	INTEGER	REPORTING PERIOD IDENTIFIER
REPORTNAME	CHAR(30)	REPORTING PERIOD NAME

BEGINDATE	DATE
ENDDATE	DATE
BEGINFILINGDATE	DATE
ENDFILINGDATE	DATE

### **CECPERIODS**

CECPeriod lookup (reportdate)

/\* Table: CECPERIODS \*/

```
CREATE TABLE CECPERIODS (
    REPORTDATE      DATE          POTENTIAL REPORTING DATES FOR CANDIDATES
```

### **ELECTIONDATES**

ELECTIONDATES define (CYCLE)

/\* Table: ELECTIONDATES \*/

```
CREATE TABLE ELECTIONDATES (
    CYCLE           INTEGER      CAMPAIGN CYCLE (yyyy)
    PRIMARYDATE    DATE         DATE OF PRIMARY FOR CYCLE
    GENERALDATE    DATE         DATE OF GENERAL FOR CYCLE
```

### **TRANSACT**

Transactions for a particular cycle, period and filerid

```
TRANSACT(FILERID)
    → FILINGS(CYCLE,PERIODNUMBER,FILERID)
```

/\* Table: TRANSACT \*/

```
CREATE TABLE TRANSACT
    FILERID          INTEGER      POINTER TO ORG TABLE
    TRANSID          INTEGER      UNIQUE IDENTIFIER
    TRANSTS          DATE         TRANSACTION RECORD TIME STAMP
    MASTERTRANSID   INTEGER      IF EXISTS, OTHERWISE TRANSID
    TRANSDATE        DATE         TRANSACTION DATE
    AMOUNT           DOUBLE       DOLLARS
    NAMEID           INTEGER      POINTER TO NAMELIST
    CREDITACCOUNTID INTEGER      CREDIT ACCOUNT (POINTER TO ACCOUNT TYPE)
    DEBITACCOUNTID  INTEGER      DEBIT ACCOUNT (POINTER TO ACCOUNT TYPE)
    BENEFIT          CHAR(1)
    CANDIDATEID     INTEGER      IF EXISTS, POINTER TO CANDIDATE
    REFNO            VARCHAR(10)  OPTIONAL
    RECEIPTNO       VARCHAR(10)
    PAID             VARCHAR(1)
    EVENTCODE       VARCHAR(10)
    EVENTDATE       DATE
    TRANSMEMO       VARCHAR(70)
    CECSOLICITORNAMEID INTEGER    POINTER TO NAMELIST OF SOLICITOR
    DEPOSITBATCH    VARCHAR(10)  OPTIONAL
    MODIFIED_DATE   DATE         LAST MODIFIED DATE
    MODIFIED_USER   CHAR(20)     LAST MODIFIED BY USER
    CLEARED         VARCHAR(1)   TRANSACTION CLEARED
    DELETED         VARCHAR(1)   TRANSACTION DELETED
```

## ACCOUNTS

Different accounts related to the transactions

ACCOUNTS(ACCOUNTID)

→ TRANSACT(CREDITACCOUNTID and DEBITACCOUNTID)

/\* Table: ACCOUNTS \*/

ACCOUNTID	INTEGER	UNIQUE IDENTIFIER
ACCOUNTNAME	CHAR(50)	
ACCOUNTTYPEID	INTEGER	POINTER TO ACCOUNT TYPE (ACCTTYPE TABLE)

## ACCTTYPE

Type of accounts

ACCTTYPE(ACCOUNTTYPEID)

→ ACCOUNTS(ACCOUNTTYPEID)

/\* Table: ACCTTYPE \*/

CREATE TABLE ACCTTYPE (

ACCOUNTTYPEID	INTEGER	UNIQUE IDENTIFIER
ACCOUNTTYPE	CHAR(20)	
BALANCESHEET	CHAR(1)	Y or N for Detail Summary Sheet needs generation

## LIMITS

Limits defined in law for running a campaign

LIMITS(CYCLE)

→ TRANSACT(CYCLE)

/\* Table: LIMITS \*/

CREATE TABLE LIMITS (

CYCLE	INTEGER	UNIQUE IDENTIFIER
CONTRIBUTIONTYPE	INTEGER	UNIQUE WITHIN CYCLE
DESCRIPTION	CHAR(90)	
REVISEDDATE	DATE	DATE LAST REVISED BY OFFICE
LOCALLIMIT	DOUBLE	APPLIES TO LOCAL JURISDICTIONS
STATEWIDELIMIT	DOUBLE	APPLIES TO STATEWIDE JURISDICTIONS
STATUTES	CHAR(60)	DEFINING STATUTE
LEGLIMIT	DOUBLE	APPLIES TO LEGISLATIVE OFFICES
CCECLEGLIMIT	DOUBLE	APPLIES TO PARTICIPATING LEGISLATIVE OFFICES
CCECSTATEWIDELIMIT	DOUBLE	APPLIES TO PARTICIPATING STATEWIDE OFFICES

## NAMELIST

Name related to Transaction for a particular cycle, period, filerid and transaction

NAMELIST(CYCLE,REPORTNUMBER,FILERID,NAMEID)

→ TRANSACT(CYCLE,REPORTNUMBER,FILERID,NAMEID)

/\* Table: NAMELIST \*/

CREATE TABLE NAMELIST

FILERID	INTEGER	POINTER TO ORGS
NAMEID	INTEGER	UNIQUE IDENTIFIER
NAMETS	DATE	TIME STAMP FOR NAME RECORD
NAMETYPEID	INTEGER	POINTER TO NAME TYPES
NAME	CHAR(50)	

ADDRESS1	CHAR(40)	
ADDRESS2	CHAR(40)	
CITY	CHAR(20)	
STATE	CHAR(2)	
ZIP	CHAR(10)	
OCCUPATION	CHAR(50)	
EMPLOYER	CHAR(50)	
PACID	INTEGER	POINTER TO ASSOCIATED PAC
FAXPHONE	VARCHAR(14)	
EMAIL	VARCHAR(30)	
USER1	VARCHAR(10)	
USER2	VARCHAR(10)	
USER3	VARCHAR(10)	
USER4	VARCHAR(10)	
AFFILIATION	VARCHAR(15)	IF EXISTS
OFFICEID	INTEGER	OFFICE LOOKUP
COUNTYID	INTEGER	COUNTY LOOKUP
MODIFIED_DATE	DATE	DATE OF MODIFICATION
MODIFIED_USER	CHAR(20)	MODIFICATION USER
DELETED	CHAR(1)	NAME DELETED

### **NAMETYPE**

Type of name record

NAMETYPE(NAMETYPEID)

→ NAMELIST(NAMETYPEID)

/\* Table: NAMETYPE \*/

```
CREATE TABLE NAMETYPE
  NAMETYPEID      INTEGER      UNIQUE IDENTIFIER
  NAMETYPE        VARCHAR(25)
```

### **COUNTIES**

County lookup table

COUNTIES(CODE)

→ ORGS(CANDCOUNTY)

/\* Table: COUNTIES \*/

```
CREATE TABLE COUNTIES
  COUNTYID        INTEGER      UNIQE IDENTIFIER
  STATE           CHAR(2)      STATE IDENTIFIER
  COUNTYNAME      VARCHAR(30)  COUNTY NAME
```

### **OFFCODE**

Office code lookup table

OFFCODE(CODE)

-> ORGS(CANDOFFICE)

/\* Table: OFFCODE \*/

```
CREATE TABLE OFFCODE
  OFFICEID        INTEGER      UNIQUE IDENTIFIER
  OFFICENAME      VARCHAR(45)  OFFICE NAME
```

