

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

**BRIEF OF *AMICI CURIAE*
FIRST AMENDMENT COALITION OF ARIZONA, INC., SOCIETY OF
PROFESSIONAL JOURNALISTS, AND ARIZONA NEWSPAPERS
ASSOCIATION**

Daniel C. Barr (010149)
K. McKay Worthington (018703)
Christopher M. Schultz (019989)
Elizabeth J. Kruschek (025081)
PERKINS COIE BROWN & BAIN P.A.
2901 N. Central Avenue, Suite 2000
Phoenix, Arizona 85012-2788
(602) 351-8000
Email: doCKETphx@perkinscoie.com

Attorneys for *Amici Curiae* First
Amendment Coalition of Arizona, Inc.,
Society of Professional Journalists, and
Arizona Newspapers Association

July 16, 2009

TABLE OF CONTENTS

	Page
Table of Authorities	ii
Interest of Amici	1
Introduction.....	2
Argument	2
I. “Metadata” Is A Broad Term, Encompassing Various Types Of Information.....	4
II. Any Concerns About Confidentiality Or Privilege In Producing Metadata Can Be Addressed Under The Current Public Records Law.....	10
III. Treating Metadata As A Public Record Provides Significant Value To Journalists And To The Public.....	13
Conclusion	15
Certificate of Compliance.....	17
Certificate of Service	18

TABLE OF AUTHORITIES

	Page(s)
Cases	
<i>Williams v. Sprint/United Mgmt Co.</i> , 230 F.R.D. 640 (D. Kan. 2005)	5, 6, 10
Regulations and Rules	
Fed. R. Civ. P. 26(f), cmt. to 2006 amends.	4
Other Authorities	
Marjorie A. Shields, <i>Discoverability of Metadata</i> , 29 A.L.R. 6th 167 (2007).....	5
The Sedona Conference, <i>The Sedona Guidelines: Best Practice Guidelines & Commentary for Managing Information and Records in the Electronic Age</i> , (<i>Second Edition</i>) (2007), http://www.thosedonaconference.org/ content/miscFiles/publications_html	4, 5, 6
The Sedona Conference, <i>The Sedona Conference Glossary: E-Discovery & Digital Information Management</i> (<i>Second Edition</i>) (2007), http://www.thosedonaconference.org/ content/miscFiles/publications_html	5

Interest of *Amici*

Amicus curiae First Amendment Coalition of Arizona, Inc. (the “Coalition”) is a nonprofit Arizona corporation formed for the purposes of educating, informing and advising the public and the news media with respect to the freedoms of speech and press under the Arizona and federal constitutions and the rights of public access granted by state and federal laws. Since 1982, the Coalition has engaged in litigation on behalf of its members to protect such freedoms and rights. The members of the Coalition include the Arizona Broadcasters Association, the Arizona Newspapers Association, the Arizona-New Mexico Cable Communications Association, the Society of Professional Journalists and the Arizona Press Club.

Amicus curiae the Society of Professional Journalists (“SPJ”) is dedicated to improving and protecting journalism. It is the nation’s largest and most broadly-based journalism organization, dedicated to encouraging the free practice of journalism and stimulating high standards of ethical behavior. Founded in 1909 as Sigma Delta Chi, the SPJ promotes the free flow of information vital to a well-informed citizenry; works to inspire and educate the next generation of journalists; and protects First Amendment guarantees of freedom of speech and press.

Amicus curiae the Arizona Newspapers Association (“ANA”) is a nonprofit trade association that represents 124 newspapers throughout Arizona. Established

in 1930, the ANA is the successor to the Arizona Press Association of 1905 and the Arizona Daily Newspaper Association of 1922.

Introduction

This case requires the Court to consider whether metadata contained within an electronic public record is subject to disclosure under the Arizona Public Records Law. Put another way, should electronic documents in their native format be treated any differently by the Public Records Law than paper documents or electronic documents in non-native formats? This amicus brief and its exhibits first explain and provide examples of how metadata has been defined, what metadata is, and how the types of available metadata vary depending on the form of the document at issue. Next, the brief explains that concerns about confidentiality or privilege in producing metadata are misplaced, because all such concerns can be addressed under the current Public Records Law. Finally, the brief provides examples of how journalists have used metadata to produce countless reports of vital public interest—reports that could not exist without access to metadata.

Argument

As in the private sector, government business is increasingly conducted using computers and electronically created documents. And, as in the private sector, the use of electronically created and maintained documents in government

business is invaluable for efficient decision-making. However, the almost ubiquitous use of electronically created and maintained documents by government agencies also means that more and more government decisions are not fully reflected in traditional hard copy documents. Although some public bodies responding to a public records request may produce requested documents in an electronic form, such as Tagged Image File Format (“TIFF”) or Portable Document Format (“PDF”), those forms of production are essentially electronic photocopies, *i.e.*, they are static images of documents, and do not contain critical information about the document itself.

Accordingly, the Appellant in this case is asking this Court to rule that the term “public records” encompasses electronic versions of documents that contain “metadata” about those documents. In essence, Appellant is arguing that public entities should produce public records in “native format” if requested to do so. Unlike traditional hard copy documents, documents produced in “native format” can be opened, viewed, and manipulated using the software application in which they were originally created, *i.e.*, they can be opened, viewed, and manipulated “natively.” Thus, although a Microsoft Word document can be scanned and produced electronically by a public entity as a static TIFF or PDF image, that document cannot be opened *as* a Word document in Word by the party making the public records request. If, however, the Word document is produced in its native

format, the requesting party will be able to open the document in Word and it will have the same appearance and characteristics that it had when originally created.

The same is true of other types of documents: if, for example, the public entity uses standard Microsoft Office applications, then native emails can be opened and examined in Outlook, native spreadsheets can be opened and examined in Excel, and so forth. Documents that are produced in this native format carry additional data, known as “metadata,” that may or may not be apparent on the face of the document.

I. “METADATA” IS A BROAD TERM, ENCOMPASSING VARIOUS TYPES OF INFORMATION.

“Metadata” is an overarching term that applies to various concepts. The unifying principle of metadata, however, is that it is comprised of information that cannot be accessed in a traditional hard copy document. At its heart, metadata is simply “data about data.” The Sedona Conference, *The Sedona Guidelines: Best Practice Guidelines & Commentary for Managing Information and Records in the Electronic Age (Second Edition)* at 28 (2007), available at http://www.thesedonaconference.org/content/miscFiles/publications_html [hereinafter “The Sedona Guidelines”]. Metadata has been variously defined as:

- “[i]nformation describing the history, tracking, or management of an electronic file,” Fed. R. Civ. P. 26(f), cmt. to 2006 amends.;

- “[d]ata typically stored electronically that describes the characteristics of [electronically stored information], found in different places in different forms. Can be supplied by applications, users, or the file system. Metadata can describe how, when and by whom [electronically stored information] was collected, created, accessed, modified, and how it is formatted,” The Sedona Conference, *The Sedona Conference Glossary: E-Discovery & Digital Information Management (Second Edition)* at 33 (2007), available at http://www.thesedonaconference.org/content/miscFiles/publications_html;
- “contextual, processing, and use information which can assist with identification and certification of scope, authenticity, and integrity of electronically stored information, including electronically stored records,” The Sedona Guidelines, at 29; and
- “hidden data that usually can only be seen when a digital document is viewed in its native format using the program that originally produced the document.” Marjorie A. Shields, *Discoverability of Metadata*, 29 A.L.R. 6th 167 (2007).

As noted above, metadata takes many forms and it “varies with different applications.” *Williams v. Sprint/United Mgmt Co.*, 230 F.R.D. 640, 647 (D. Kan. 2005). “As a general rule of thumb, the more interactive the application, the more

important the metadata is to understanding the application's output." *Id.* Metadata can include information *about* a particular document, as well as information embedded *within* a particular document. The Sedona Guidelines, at 28; *see also Williams*, 230 F.R.D. at 646. Metadata *about* a particular document includes such information as the author of the file, the date the file was created, when and by whom the file was last modified, the size of the file, the location or "path" of storage for the file, and the filename. The Sedona Guidelines, at 28; *see also Williams*, 230 F.R.D. at 646.

Metadata *within* the particular document is not apparent from the face of the document. The classic example of this type of metadata is the hidden tracked changes in a Microsoft Word document. If a native Word document contains tracked changes, those changes may not be visible when the file is initially opened, but they can be viewed simply by clicking the "Track Changes" button in Word.

Exhibit A of this brief, which contains PDF and native Word versions of the same hypothetical letter, illustrates the importance of the "Track Changes" feature in the public records context. The PDF version of the document simply shows that a governmental agency rejected the contention that one of its members should have recused himself from consideration of a requested zoning variance because of a conflict of interest and informs the interested party that the variance has been denied. In comparison, the native Word document shows that the governmental

agency not only changed its position regarding the potential conflict of interest prior to sending out the letter, but also backdated the letter. Without access to metadata, this critical fact would not come to light.

Numerous other forms of metadata are embedded or “hidden” within particular documents. Microsoft Outlook emails, for example, contain details about the sender’s domain, the route on the Internet over which the message was transmitted, and any delays that may have occurred during transmission.¹ Additionally, the formulae behind the numbers in spreadsheets, comments inserted into Word documents and/or PowerPoint presentations, hidden worksheets within a spreadsheet, and hidden slides or speaker’s notes within a PowerPoint presentation are all forms of metadata.

Exhibit A also provides an example of hidden metadata in a hypothetical Excel spreadsheet. If the spreadsheet is opened as a PDF (or viewed in hard copy form), there are only three columns showing the zoning variance at issue, the date of the vote on that variance, and the disposition of the variance. The native spreadsheet, in contrast, contains additional, hidden data that sheds light on the votes of individual agency members and discussions about potential recusals by

¹ Although the “To,” “From,” “Subject,” “Date,” “Time,” “cc,” and “bcc” fields in emails are sometimes characterized as metadata, this information is often apparent on the face of a printed email and therefore is not metadata in the traditional sense of the word.

individual members. Once again, without the ability to access the metadata, this information would not be uncovered by a public records request.

In addition to the types of metadata discussed above, and as set forth in the accompanying Declaration of Stephen K. Doig (“Doig Declaration,” attached as Ex. B), coded datasets created and maintained by government agencies to track anything from demographic changes to the number of accidents at particular intersections—information that is invaluable to journalists, who frequently make public records requests for just this type of information—contain additional types of metadata. [Ex. B ¶¶ 5-9] Although this type of metadata is not specifically at issue in this case, it is critical to journalists working with government datasets and it will be covered by any opinion this Court issues on the availability of metadata in the public records context.

Examples of the types of metadata associated with such datasets include the file layout of the dataset, which explains how to parse each line of the data into proper variables; the data dictionary, which explains the codes and abbreviations necessary to understanding the meaning of the variable; and the relational architecture, which allows the user to fully use the dataset as a relational database so that all variables can be properly analyzed. [*Id.* ¶ 5. The Appendix to the Doig Declaration contains the “Campaign Finance Database Table Specification,” an example of the metadata associated with just one dataset.]

As a simple example, following are two hypothetical records from a government-maintained computer dataset:

194804211722388500410

195008030691728528121

[*Id.* ¶ 7] The metadata, which permits interpretation of these strings of numbers, is shown below and combines the file layout (the “Column” heading, which explains how to parse the numbers into meaningful variables) and the data dictionary (the “Variable” heading, which explains the meaning of each variable):

<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)

[*Id.* ¶ 8] Thus, the first four numbers or “columns” in each of the datasets, 1948 in the first instance and 1950 in the second, correspond to the birth year of the hypothetical individual at issue. The fifth and sixth numbers or “columns” in each of the datasets corresponds to the birth month of those individuals, and so forth.

Applying the metadata to the two hypothetical records, we learn that the first set of numbers describes a White male who was born on April 21, 1948, is 6 feet tall, weighs 238 pounds, and lives in the downtown Phoenix zip code of 85004. The second hypothetical record describes a black Hispanic female who was born

on August 3, 1950, is 5 feet, 9 inches tall, weighs 172 pounds, and lives in the Tempe zip code of 85281. [*Id.* ¶ 9]

As is clear from even the simple examples above, production of the metadata associated with these datasets is critical. Unlike Word documents and Outlook emails, which can be read without metadata, the records in these datasets *cannot* be interpreted or used without the associated metadata. *See Williams*, 230 F.R.D. at 647 (“The metadata is the key to showing the relationships between the data [in a database]; without such metadata, the tables of data would have little meaning.”). Producing the dataset without the metadata required to interpret it is analogous to releasing a password-protected file without the password. In short, permitting a public entity to refuse to produce a dataset’s metadata allows the public entity to effectively deny a public records request. [Ex. B ¶ 10]

II. ANY CONCERNS ABOUT CONFIDENTIALITY OR PRIVILEGE IN PRODUCING METADATA CAN BE ADDRESSED UNDER THE CURRENT PUBLIC RECORDS LAW.

In its Supplemental Brief (at 14-15), the City of Phoenix suggests that production of metadata under the Public Records Law would “require an entity to exhaust countless hours and incur the added expense of computer experts just to determine what metadata exists.” The City exaggerates. First, this case concerns a very narrow request for specific metadata associated with electronic notes, the production of which the City cannot seriously argue would be unduly burdensome.

