I/S: A JOURNAL OF LAW AND POLICY FOR THE INFORMATION SOCIETY

More Than Just Ones and Zeros: The Reproducibility of Metadata Under the Freedom of Information Act

CHRISTOPHER R. MELTZER*

I. INTRODUCTION

Suppose you believe that the government is misrepresenting the benefits of a newly implemented program created by the Department of Homeland Security (DHS) regarding immigration and border protection. In your attempt to unveil the truth, you file a request under the Freedom of Information Act (FOIA)[1] for documents and records circulated within the Department to be reproduced for you, electronically. Would the records you receive include the times the documents were created, who created them, a list of any changes made or comments added? Does this electronically stored information have to be produced even though you did not request it? Does it have to be produced even if you were to request it?

Currently the answer, as with most legally-posed questions, is "it depends." While you may find it necessary that the government produce such information routinely in order to adhere to the principles of FOIA, namely openness and public access to government

* J.D. Candidate, The Ohio State University Moritz College of Law, 2013; B.A. in History, *with honors*, University of California, Santa Barbara, 2009. The views expressed in this article are solely the views of the author. I would like to thank Debbie Diener for introducing me to the topic and Professor Peter Swire, without whom this Article would not have been a success. I would also like to thank Kailee Goold and Chris Hammond for their comments and support throughout the entire writing process. Finally, I would like to thank my amazing fiancée, Kaileen, for her unequivocal support, dedication, and for putting up with all of the 'law school talk' throughout the past few years.

[1] 5 U.S.C. § 552 (2006).

records, [2] suppose, however, that the person requesting the information was not you, rather it was a terrorist or other adversary to the United States.[3] Suppose also that some information in the records was redacted[4] from the documents before disclosure because it was information related to national security.[5] Would you still want to require the government to routinely produce the records in native format, [6] risking that any information previously redacted be discovered by the adversary?[7]

---

[2] *See* U.S. DEP'T OF JUSTICE, THE UNITED STATES DEPARTMENT OF JUSTICE GUIDE TO THE FREEDOM OF INFORMATION ACT 1 (2009) [hereinafter DOJ GUIDE TO FOIA] (citations omitted) (internal quotation marks omitted) ("[T]he basic purpose of the FOIA is to ensure an informed citizenry . . . . FOIA is often explained as a means for citizens to know what their Government is up to.").

[3] After all, "FOIA does not permit selective disclosure of information only to certain parties." P. STEPHEN GIDIERE III, THE FEDERAL INFORMATION MANUAL 353 (2006).

[4] Redaction is "the rendering of data so that it is unreadable or is truncated so that no more than the last four digits of the identification number is accessible as part of the data." *See* Mintz, Levin, Cohn, Ferris, Glovsky and Popeo, P.C., *State Data Security Breach Legislation Survey*, 37-20 THE LAWYER'S BRIEF II 26 (2007), *available at* http://www.mintz.com/newsletter/2007/PrivSec-Alert-DataBreachLaws-08-07/state_data_breach_matrix_0807.pdf.

[5] 5 U.S.C. § 552(b)(1); *see also* DOJ GUIDE TO FOIA, *supra* note 2, at 147.

[6] A native file is a document that has "an associated file structure defined by the original creating application." *See* THE SEDONA CONFERENCE WORKING GROUP SERIES, THE SEDONA CONFERENCE GLOSSARY: E-DISCOVERY & DIGITAL INFORMATION MANAGEMENT 35 (3d ed. 2010), *available at* https://thesedonaconference.org//publication/The%2520Sedona%2520Conference%25C2%25AE%2520Glossary [hereinafter THE SEDONA CONFERENCE GLOSSARY]; GEORGE L. PAUL & BRUCE H. NEARON, THE DISCOVERY REVOLUTION 98 (2005) ("Native format is the default format of the software that was used to create a file . . . . For example, the default native format of a spreadsheet file generated by Microsoft Excel will be in the 'xls' format. A default native format of a word-processing document generated by Microsoft Word will be the 'doc' format."); *see also* Lake v. City of Phoenix, 218 P.3d 1004, 1008 (Ariz. 2009) (holding that an agency "can satisfy a public records request merely by providing the requestor with a copy of the record in its native format").

[7] Information redacted from the native version of a file is never fully deleted from the file and can be recovered with the appropriate forensic computer skills. *See* DAVID L. MASTERS, THE LAWYER'S GUIDE TO ADOBE ACROBAT 213-14 (2008). In addition, un-redacting a document can be relatively easy, especially when the document is in its native format. For more information on the un-redacting of documents see Ari Kaplan, *Redact the Right Way: Text Hidden from View and Not Seen in Printed Format can be Recovered if the Document is Submitted Electronically*, N.J. L.J., Feb. 10, 2003, at 66 ("Word retains hidden information within its documents to allow users to undo mistakes, resulting in security problems when electronically filing a Word document."); *see also*, ARCHITECTURES AND APPLICATIONS DIVISION OF THE SYSTEMS AND NETWORK ATTACK CENTER (SNAC)

When put into context, the decision whether to mandate government production of records in native format becomes much more difficult—on the one hand is the public's interest in maintaining an open and transparent government and on the other is the need to safeguard certain sensitive information.[8] As a result, this Article seeks to balance these two opposing ideals by providing an explanation of what metadata is and proposing a framework of when and how metadata should be reproduced. [9] State courts are now being confronted with an increasing amount of litigation regarding whether metadata needs to be produced under state public records laws.[10] However, the issue of whether metadata needs to be produced under federal FOIA has never officially been addressed. As a result, this

---

INFORMATION ASSURANCE DIRECTORATE, NATIONAL SECURITY AGENCY, REDACTING WITH CONFIDENCE: HOW TO SAFELY PUBLISH SANITIZED REPORTS CONVERTED FROM WORD TO PDF (2005).

[8] "A democracy requires accountability, and accountability requires transparency. The FOIA encourages accountability through transparency." *See* DOJ GUIDE TO FOIA, *supra* note 2 at 20 (citing Presidential Memorandum for Heads of Executive Departments and Agencies Concerning the Freedom of Information Act, 74 Fed. Reg. 4683 (Jan. 21, 2009)); *cf.* John Doe Agency v. John Doe Corp., 493 U.S. 146, 152 (quoting H.R. REP. NO. 89-1497, at 6 (1966), *reprinted in* 1966 U.S.C.C.A.N. 2418, 2423) where the Supreme Court held that "Congress sought to 'reach a workable balance between the right of the public to know and the need of the Government'" to withhold sensitive information.

[9] *See infra* Parts II & V.

[10] The only case to have addressed the issue of metadata production under federal FOIA was Nat'l Day Laborer's Org. Network v. U.S. Immigration & Customs Enforcement, No. 10 Civ. 3488 (SAS) document 41 (S.D.N.Y. 2011), *available at* http://ccrjustice.org/files/Doc%2041%202-7-11%20Opinion%20and%20Order%20re%20Form%20of%20Production.pdf. The opinion regarding form of production, however, has since been withdrawn and reconsidered on other grounds. *See id.* at document 98, *available at* http://ccrjustice.org/files/6-17-11%20Order%20Withdrawing%202-14-11%20Format%20of%20Production%20Order.pdf; *cf.* Lake v. City of Phoenix, 218 P.3d 1004, 1007 (Ariz. 2009) (en banc) ("The metadata in an electronic document is part of the underlying document; it does not stand on its own. When a public officer uses a computer to make a public record, the metadata forms part of the document as much as the words on the page."); O'Neill v. City of Shoreline, 240 P.3d 1149, 1154 (Wash. 2010) (en banc) ("We agree with the Supreme Court of Arizona that an electronic version of a record, including its embedded metadata, is a public record subject to disclosure."); *but see* Irwin v. Onondaga Cnty. Res. Recovery Agency, 72 A.D.3d 314, 319, 895 N.Y.S.2d 262, 266 (N.Y. App. Div.2010), where the court limited their decision of metadata production to the limited facts of the case as the area of law is still evolving, noting that "[t]he issue of whether metadata is subject to disclosure has been broached in a number of other jurisdictions, and we consider informative but not dispositive the decision of the Supreme Court of Arizona in *Lake v. City of Phoenix*."

Article provides the first in-depth analysis regarding metadata production under the Freedom of Information Act.

In order to fully comprehend the nature of the issue, it is necessary to first understand what metadata is and how it has been incorporated into ordinary judicial proceedings. As such, Part II of this Article defines what metadata is, how it has been treated in ordinary litigation, and explores why metadata is becoming increasingly important in the legal field. Though it has been held that metadata is generally discoverable in ordinary litigation, the difference between requests under FOIA and the reproducibility of metadata under civil discovery rules is significant. As a result, while the Federal Rules of Civil Procedure (FRCP) could seemingly provide a framework for how and when metadata should be reproduced, Part III distinguishes between the FRCP and FOIA and explains how producing records in native format, while acceptable in ordinary litigation, should not be accepted under FOIA.[11]

Since there is currently no framework in place, the question arises of whether, and to what extent, metadata should have to be produced when requested under FOIA. To this end, there is no general rule applicable to all types of metadata. Some argue that the native version of a public record including corresponding metadata should be furnished in all cases in which it is requested.[12] However, Part IV explains that because of the complexities of the Freedom of Information Act and the potential for severe consequences, to create a general rule applicable to all requests of metadata under FOIA is impracticable. As a result, Part V of this Article proposes a framework where records produced under FOIA should neither presumptively include metadata nor be produced in native format. Instead, when metadata is specifically requested under FOIA, the agency should reproduce the record with corresponding metadata only if the

---

[11] The court's decision in *Nat'l Day Laborer*, No. 10 Civ. 3488 (SAS) document 41, based the form of metadata production under FOIA on the Federal Rules of Civil Procedure, an analysis that is inherently flawed. While the opinion has been withdrawn and serves no precedential value, the case presents a framework that a court may use in the future. This Article hopes to prevent that framework from being utilized by an independent court in the future.

[12] Writing in response to state public records laws, *see* Peter S. Kozinets, *Access to Metadata in Public Records: Ensuring Open Government in the Information Age*, COMM. LAWYER, July 2010, at 1 ("[M]ost metadata can only be seen when viewing an electronic record in its native format . . . and protecting public access to electronic records, including metadata, is essential to safeguarding the public's ability to open government conduct to public scrutiny.").

metadata is not exempt and is readily reproducible. If the metadata requested satisfies these requirements, it should be produced only in static image format with corresponding load files.[13] As this proposed framework properly balances the two opposing concepts of promoting government transparency and safeguarding sensitive information, it needs to be considered before any decision is rendered that mandates the production of records in a format that can cause a serious threat to the nation's security or individuals' privacy.

II. WHAT IS METADATA?

Frequently described as "data about data," metadata is electronically-stored data that describes the characteristics of electronically-stored information (ESI) "such as how, when, and by whom the ESI is collected, created, accessed, modified, and how it is formatted."[14] Metadata can be created by applications, users, or the file system and can be altered intentionally or inadvertently.[15] Essentially, metadata "record[s] information about the document or file automatically to assist [the user] in storing and retrieving the

---

[13] A load file is a separate file that may "contain electronic text and metadata to accompany the [static] images." Jeffrey Gross, *Objection to Form: Rule 34(b) and the Form of Production of Electronically Stored Information*, THE PRAC. LITIGATOR, July 2009, at 39, 41. See also Aguilar v. Immigration & Customs Enforcement Div. of U.S. Dep't of Homeland Sec., 255 F.R.D. 350, 353 n. 3 (S.D.N.Y. 2008) (citing PSEG Power N.Y., Inc. v. Alberici Constructors, Inc., No. 1:05-CV-657 (DNH)(RFT), 2007 WL 2687670, at *2 n. 2 (N.D.N.Y. Sept. 7, 2007)), explaining that static image format is a TIFF or PDF file "that creates a mirror image of the electronic document"; *In re* Payment Card Interchange Fee & Merch, Disc., No. MD 05-1720(JG)(JO), 2007 WL 121426, at *1 n. 2 (E.D.N.Y. Jan. 12, 2007).

[14] THE SEDONA CONFERENCE GLOSSARY, *supra* note 6, at 34. *See also* PAUL & NEARON, *supra* note 6, at 40.

[15] W. Lawrence Wescott II, *The Increasing Importance of Metadata in Electronic Discovery*, 14 RICH. J.L. & TECH. 10, 15 (2008). For example, when a document is created on a computer, the 'author' of the document will be assigned based on the current owner of the computer. However, when that file is forwarded electronically to another individual for editing, the 'author' will remain as the owner of the computer that the document was originally created on. This is also true for document templates that may be used routinely by various members of an agency. When a template is created, the 'author' of the document will be the computer owner. Various changes and modifications to the template after it has been distributed could all potentially be associated with that original computer owner if not manually changed. This could lead to a document having as its 'author' a person who has never before seen or touched the final document. *Id.*

document or file at a later date."[16] In addition, metadata is generally not reproduced in full form when a document is printed to paper or reproduced as a static, electronic image.[17]

A. *Organizing Metadata into Three General Categories*

Metadata can be grouped into three broad categories: substantive metadata (also referred to as application metadata), system metadata, and embedded metadata.[18] First, substantive, or application metadata, is created by the application specific to the ESI being addressed, embedded in the file, and moved with the file when copied.[19] This data may reflect substantive changes to a document by the user and/or instruct the software program on how to display the document[20] because substantive metadata "records and reflects any changes to a document made by the user or creator of a document."[21] Examples include the track changes function in a Microsoft Word document[22] and other internal data such as who created the document, any revisions that were made, and when the revisions occurred.[23] This type of data is often of much concern to attorneys in ordinary

---

[16] *See* Mathew Robertson, *Why Invisible Electronic Data is Relevant in Today's Legal Arena*, 23 J. AM. ACAD. MATRIM. LAW. 199, 201 (2010); Wescott, *supra* note 15, at 3.

[17] *See also* PAUL & NEARON, *supra* note 6, at 105; Steven C. Bennett & Jeremy Cloud, *Coping with Metadata: Ten Key Steps*, 61 MERCER L. REV. 471, 471 (2010). For examples and further explanation, *see* Robert L. Kelly, *The Tech Side of E-Discovery: Understanding Electronically Stored Information*, BUS. LAW TODAY, Oct. 17, 2007, at 43, 45–46.

[18] See *Aguilar*, 255 F.R.D. at 354. The varying types of metadata have been separated into categories for sake of clarification. While the distinction is important and relevant to the framework presented in Part V, it is not necessary for the reader to become an expert in distinguishing between the types to understand the concept behind this Article.

[19] Jay E. Grenig & William C. Gleisner, III, *Metadata*, *in* EDISCOVERY & DIGITAL EVIDENCE §1.5 (2011); *Aguilar*, 255 F.R.D. at 354; THE SEDONA CONFERENCE GLOSSARY, *supra* note 6, at 3.

[20] THE SEDONA CONFERENCE WORKING GROUP SERIES, THE SEDONA PRINCIPLES: BEST PRACTICES RECOMMENDATIONS & PRINCIPLES FOR ADDRESSING ELECTRONIC DOCUMENT PRODUCTION 60 (2d ed. 2007) [hereinafter THE SEDONA PRINCIPLES].

[21] Robertson, *supra* note 16, at 202.

[22] For an example of Track Changes, *see* PAUL & NEARON, *supra* note 6, at 105.

[23] Robertson, *supra* note 16, at 203; Wescott, *supra* note 15, at 3–4.

litigation, as such data has the potential to reveal important and confidential information[24] if inadvertently sent by an attorney or uncovered by opposing counsel.[25] Substantive metadata is also prone to a large amount of contextual inaccuracy.[26]

Second, embedded metadata is generally hidden but usually considered to be an integral part of ESI.[27] This data is embedded in a file and is only available in the original, native file.[28] It consists of text, numbers, hyperlinks, data, or any other information that is not observable by an individual "viewing the output display of the native file."[29] An example of embedded metadata is that of an Excel spreadsheet that uses formulas that underlie the output of a cell—the formulas underlying the document would constitute metadata.[30]

---

[24] Such information could consist of personally identifiable information of the client or work product between the client and attorney. *See* David K. Isom, *Electronic Discovery Primer for Judges*, 2005 FED. CTS. L. REV. 1, [II.O.10–11] (2005), *available at* http://www.fclr.org/fclr/articles/html/2005/fedctslrev1.shtml; *see also infra* Parts IV.B & IV.C.

[25] *See* Robertson, *supra* note 16, at 203 ("The 'track changes' function . . . shows any alterations made to previous drafts of a document and the identities of the users who made the changes . . . . Though the 'tracked changes' [may be] deleted on the viewable surface of a document, the 'tracked changes' are often still stored within the substantive metadata . . . [and] if it is not removed, [it can] reveal secret information to other parties.").

[26] *See infra* Part IV.C.

[27] Grenig & Gleisner, *supra* note 19; Aguilar v. Immigration & Customs Enforcement Div. of U.S. Dep't of Homeland Sec., 255 F.R.D. 350, 354 (S.D.N.Y. 2008); THE SEDONA CONFERENCE GLOSSARY, *supra* note 6, at 19.

[28] When a file is converted to static-image format, the embedded metadata will typically be lost. *See* Jason Krause, *Sloppy Redaction: To Err is Automated*, N.J. L.J., Aug. 20, 2009, at 26.

[29] *See* Wescott, *supra* note 15, at 4,

> Examples include: spreadsheet formulas (which display as the result of the formula operation), hidden columns, externally or internally linked files (*e.g.*, sound files in Powerpoint presentations), references to external files and content (*e.g.*, hyperlinks to HTML files or URLs), references and fields (*e.g.*, the field codes for an auto-numbered document), and certain database information if the data is part of a database (*e.g.*, a date field in a database will display as a formatted date, but its actual value is typically a long integer)

See also Robertson, *supra* note 16, at 204.

Without the ability to view these formulas, the spreadsheet could be incomprehensible and may not provide any beneficial use to the requesting party.[31]

Third, system metadata is created automatically by the operating system to track the demographics of ESI.[32] This data includes logs and other logistical information generated by the operating system to track modifications of a record's name, size, location, and the date and time of creation.[33] System metadata is generally useful in determining the authenticity of a document as it could reveal "information regarding the identity of the author and the date and time of creation . . . [and] is created automatically by the user's application or operating system."[34]

---

[30] *See* Williams v. Sprint/United Mgmt. Co., 230 F.R.D. 640, 647 (D. Kan. 2005) ("Microsoft Excel spreadsheets . . . [contain] author name or initials . . . hidden text or cells, template information, other file properties and summary information, non-visible portions or embedded objects, personalized views, and comments.").

[31] *See* Robertson, *supra* note 16, at 204 ("Spreadsheet and database output often contain calculations, query formulas, or hidden columns that are not visible in printed versions and can only be accessed within the 'native' applications . . . . [Therefore,] spreadsheet output may be difficult to understand without the ability to view the formulas underlying the printed output."); *see also Aguilar*, 255 F.R.D. at 355 ("[A] complicated spreadsheet may be difficult to comprehend without the ability to view the formulas underlying the output in each cell.").

[32] Grenig & Gleisner, *supra* note 19; *Aguilar*, 255 F.R.D. at 354; THE SEDONA CONFERENCE GLOSSARY, *supra* note 6, at 22.

[33] *See* THE SEDONA CONFERENCE GLOSSARY, *supra* note 6, at 22; *see also* Wescott, *supra* note 15, at 2.

[34] Robertson, *supra* note 16, at 204. The production of system metadata has the least potential for devastating consequences if it was released to the public, as it typically does not store any confidential or personally identifiable information. In non-FOIA related cases, "[c]ourts have commented that most system . . . metadata lacks evidentiary value because it is not relevant." *Aguilar*, 255 F.R.D. at 354 (citing Mich. First Credit Union v. Cumis Ins. Soc'y, Inc., No. Civ. 05-74423, 2007 WL 4098213, at *2 (E.D.Mich. Nov.16, 2007); Ky. Speedway, LLC v. Nat'l Assoc. of Stock Car Auto Racing, No. Civ. 05-138, 2006 WL 5097354, at *8 (E.D.Ky. Dec.18, 2006); Wyeth v. Impax Labs., Inc., 248 F.R.D. 169, 170 (D.Del. 2006)). In addition, courts have generally only found system metadata relevant if the requesting party is trying to establish "who received what information and when." *Aguilar*, 255 F.R.D. at 354. Because system metadata is "created automatically by the user's application or operating system . . . system metadata can potentially provide an objective means of authenticating many electronic documents. However, despite the date and time stamps that are automatically created . . . an individual who alters a document may not be the individual that the operating system says." Robertson, *supra* note 16, at 204; *see also* Wescott, *supra* note 15, at 22.

B. *The Increasing Importance of Metadata in the Electronic Age*

Metadata has become an increasing focus of court decisions and academics in recent years[35] because certain metadata is helpful in determining the authenticity and integrity of a record.[36] In addition, metadata is considered "a critical part of the overall functioning of a computer application . . . [b]ut metadata is not usually viewed by people looking at a screen or a printout."[37] For these reasons, the ability of an attorney to see underlying metadata of certain documents has become an important role in the discovery process and throughout litigation.[38]

In order to understand the benefits and consequences of metadata production in ordinary litigation, consider the following example. Say Defendant, *D*, is on trial for murder and her lawyer, *L*, is representing her. *L* sends a document regarding the position and location of the murder weapon after the murder via e-mail for *D* to review. *D* uses the track changes function on her computer to add a comment regarding the placement of the weapon at the crime scene. *L* receives the document and deletes the comment[39] because it has the potential to

---

[35] Grenig & Gleisner, *supra* note 19; *see also* Williams v. Sprint/United Management Co., 230 F.R.D. 640 (D. Kan. 2005); *Aguilar*, 255 F.R.D. at 354; Mike Breen, Comment, *Nothing to Hide: Why Metadata Should be Presumed Relevant*, 56 U. KAN. L. REV. 439, 447–50 (2008).

[36] *See* Kozinets, *supra* note 12, at 22 (finding that metadata "can verify the authenticity and integrity of a public record, reveal what officials knew about critical actions or decisions and when they knew it, and render intelligible vast storehouses of government data that would otherwise be useless when separated from their metadata").

[37] PAUL & NEARON, *supra* note 6, at 100. While metadata exists in every document, individual users typically do not see it as it underlies the text of a document. When viewing a document in print or on a computer screen, the user will not see the metadata unless they manually search for it. For examples of metadata created in various software programs, see Gretchen J. Harris, *Metadata: High-Tech Invisible Ink Legal Considerations*, 78 MISS. L.J. 939, 941–43 (2009); *Find and Remove Metadata (Hidden Information) in Your Legal Documents*, MICROSOFT OFFICE ONLINE, http://office.microsoft.com/en-us/help/find-and-remove-metadata-hidden-information-in-your-legal-documents-HA001077646.aspx (last visited Sept. 2, 2012).

[38] For a further elaboration on the process by which metadata can be requested and admitted into evidence in formal litigation, see Robertson, *supra* note 16, at 209–14; Harris, *supra* note 37, at 955–62.

[39] Though an individual may delete a comment while using the track changes function, it may not be permanently deleted from the document. While *L* may not see it at the time she deleted it, it is still recoverable. MASTERS, *supra* note 7, at 213–14; *see also infra* Part IV.

be construed by the jury as evidence of *D*'s guilt (by acknowledging that *D* knew where the weapon was or was not placed). Prosecutor, *P*, requests the document during discovery. If *L* were to furnish the document in PDF or TIFF format, [40] the comment would be untraceable and never discovered by *P*. However, if *L* sent the document in native format with metadata attached, then *P* could discover the comment[41] and potentially use it as evidence in trial.[42]

Because of the potential consequences of producing metadata, there is rising concern and debate among courts as to whether the documents requested by an opposing party during discovery should always be produced in native format.[43] As a result, some academics

---

[40] *See* THE SEDONA CONFERENCE GLOSSARY, *supra* note 6, at 39, 50. A PDF is a file format that captures information "from a variety of applications in such a way that they can be viewed and printed as they were intended in their original application by practically any computer, on multiple platforms, regardless of the specific application in which the original was created. PDF files may be text-searchable or image-only." *Id.* at 39. In order to view PDF files, an individual must have Adobe Reader. In addition, "Adobe® Acrobat, an application marketed by Adobe Systems, is required to edit, capture text, or otherwise manipulate a file in PDF format." *Id.*; see further *id.* at 50, describing a TIFF file as a "supported graphic file format[] for storing bit-mapped images, with many different compression formats and resolutions. File name has .TIF extension. Can be black and white, gray-scaled, or color. Images are stored in tagged fields, and programs use the tags to accept or ignore fields, depending on the application."

[41] A common way that metadata is discovered by a party is through fault of the sender. For example, when an individual clicks "final view" in Microsoft Word, all of the track changes and metadata seemingly disappear. However, when the file is sent and the metadata is not cleared, such track changes and metadata will reappear on the recipient's screen. Even if the sender were to properly remove the track changes and other metadata, a person with a higher knowledge of computer forensics may be able to recover the data anyway using advanced computer software and techniques. *See* MASTERS, *supra* note 7, at 213–14; *Find and Remove Metadata (Hidden Information) in Your Legal Documents*, *supra* note 37.

[42] One should assume for the purposes of this example that the document did not constitute work product, the attorney-client privilege did not apply, and that the metadata was not sent inadvertently thereby causing an ethical issue under the Model Code of Professional Conduct.

[43] *See, e.g.*, Breen, *supra* note 35, at 440 ("A rule that presumes the relevance of metadata better serves judicial economy and is consistent with the intent of the Federal Rules of Civil Procedure."); *see also* Williams v. Sprint/United Management Co., 230 F.R.D. 640, 656 (D. Kan. 2005) ("When the Court orders a party to produce an electronic document in the form in which it is regularly maintained, i.e., in its native format or as an active file, that production must include all metadata"); Aguilar v. Immigration & Customs Enforcement, 255 F.R.D. 350, 356 (S.D.N.Y. 2008) ("Weighing the advantages and disadvantages of different forms of production, the Conference concluded that even if native files are requested, it is sufficient to produce memoranda, emails, and electronic records in PDF or TIFF format accompanied by a load file containing searchable text and selected

have expressed their opinions on how courts should treat metadata in ordinary litigation.[44] Some states are even beginning to witness disputes over metadata production under state public records laws,[45] and it is only a matter of time before the issue is brought to the federal level. Though multiple academics and courts have discussed the concept of metadata production in ordinary litigation, what is a largely unexplored concept is how to apply these principles, despite being unclear, to the federal Freedom of Information Act.

III. ELECTRONIC DISCOVERY PROCEDURES IN CIVIL LITIGATION CANNOT BE APPLIED TO RECORDS REQUESTED UNDER FOIA

Courts in civil litigation are increasingly requiring that records be produced in native format[46] because the courts have determined that metadata is a public record and is of genuine public interest and therefore should be produced when requested.[47] In addition, metadata

---

metadata."); Wyeth v. Impax Labs., Inc., 248 F.R.D. 169, 171 (D.Del. 2006) ("Emerging standards of electronic discovery appear to articulate a general presumption against the production of metadata.").

[44] *See generally* Breen, *supra* note 35; Kozinets, *supra* note 12 (advocating that courts should adopt a rule where metadata is given a presumption of relevance). *But cf.* Douglas L. Rogers, *A Search for Balance in the Discovery of ESI Since December 1, 2006*, 14 RICH. J.L. & TECH. 8 (2008); Bennett & Cloud, *supra* note 17, at 474–75 (advocating a position that lawyers should ask questions and increase their knowledge about metadata and its importance prior to discovery in litigation).

[45] The concept of metadata production under state public records laws is beyond the parameters of this note. *But cf.* Lake v. City of Phoenix, 218 P.3d 1004 (Ariz. 2009); O'Neill v. City of Shoreline, 240 P.3d 1149 (Wash. 2010) (en banc); Irwin v. Onondaga Cnty. Res. Recovery Agency, 895 N.Y.S.2d 262 (N.Y. App. Div. 2010). For an explanation of the effects of electronic information under state public records laws, see generally ACCESS TO GOVERNMENT IN THE COMPUTER AGE: AN EXAMINATION OF STATE PUBLIC RECORDS LAWS (Martha Harrell Chumbler ed., 2007).

[46] Breen, *supra* note 35, at 439; *see also Williams*, 230 F.R.D. at 656 ("When the Court orders a party to produce an electronic document in the form in which it is regularly maintained, i.e., in its native format or as an active file, that production must include all metadata."); THE SEDONA PRINCIPLES, *supra* note 20, at 60 cmt. 3.d; Nova Measuring Instruments Ltd. v. Nanometrics, Inc., 417 F. Supp. 2d 1121, 1122 (N.D. Cal. 2006) ("So there is no confusion, if [Nanometrics] has not already done so, it must produce the documents in their native file format, with original metadata."); *In re* Verisign, 2004 WL 2445243 at *3 (N.D. Cal. 2004) (holding that producing documents in native format with metadata was "not clearly erroneous").

[47] *See supra* note 34 and accompanying text.

may constitute a public record under the Federal Records Act.[48] Nevertheless, to adopt a model that requires a record to be produced in native format so that the metadata is reproduced whenever requested under federal FOIA would not only be impracticable but may cause substantial harm to privacy interests and compelling government interests, beyond any risk associated with ordinary litigation.[49] Although producing information in native format can have evidentiary benefits to those who request information in ordinary litigation,[50] the probative value of releasing records in native format under FOIA diminishes in comparison to any genuine public interest in viewing the underlying metadata.[51]

Using the Federal Rules of Civil Procedure to govern metadata production under FOIA is inherently flawed. However, the only court to have addressed the issue of metadata production under FOIA utilized the FRCP as a framework for determining when and how metadata should be produced.[52] As a result of this confusion, this section provides a distinction between document production under the Federal Rules of Civil Procedure and document production under FOIA. It then elaborates on the actual text of FOIA to demonstrate that while the Act states that a party may request a document in "any form or format . . . if the record is readily reproducible,"[53] a federal

---

[48] Kozinets, *supra* note 12, at 24; *see also* Federal Records Act, 44 U.S.C. § 3301 (2006) ("'[R]ecords' includes all books, papers, maps, photographs, machine readable materials, or other documentary materials, regardless of physical form or characteristics.").

[49] *See* ALAN CHARLES RAUL, PRIVACY AND THE DIGITAL STATE: BALANCING PUBLIC INFORMATION AND PERSONAL PRIVACY 52–72 (2001).

[50] *See* Breen, *supra* note 35, at 440 ("Metadata gives meaning to much electronic information, and thus its presence or absence can be outcome determinative."); Robertson, *supra* note 16, at 209 ("Despite this judicial skepticism, metadata may have significant evidentiary value.").

[51] This is not to say that metadata should never be released; however, any record requested should not be released in native form. *See infra* Parts IV.B & IV.C.

[52] Nat'l Day Laborer's Org. Network v. U.S. Immigration & Customs Enforcement, No: 10 Civ. 3488 (SAS) document 41 (S.D.N.Y. 2011), *available at* http://ccrjustice.org/files/Doc%2041%202-7-11%20Opinion%20and%20Order%20re%20Form%20of%20Production.pdf. The opinion regarding form of production, however, has since been withdrawn and reconsidered on other grounds. *See id.* at document 98, *available at* http://ccrjustice.org/files/6-17-11%20Order%20Withdrawing%202-14-11%20Format%20of%20Production%20Order.pdf.

[53] 5 U.S.C. § 552(a)(3)(B)(2006).

agency should not have to produce such record(s) in native format because of the severe consequences that would result.

A. *Introducing the Freedom of Information Act (FOIA)*

The Freedom of Information Act was enacted in 1966 to promote government transparency and to ensure public access to agency records and information.[54] The Act has since been amended several times, most notably in 1996 with the Electronic Freedom of Information Act (e-FOIA) amendments.[55] The purpose of these amendments was to acknowledge the government's increased use of electronic technology and to encourage agencies to use such technology to enhance public access to government records.[56] Information available under FOIA includes all federal agency records except records or portions of records that are protected under one of the Act's nine exemptions or three exclusions.[57] An individual may, under FOIA, request and receive by mail *any federal agency's files* as long as it is not covered by one of the exemptions or exclusions.[58] For example, an individual may make a FOIA request to the Consumer Product Safety Commission for recall documents about a certain toy posing a safety hazard.[59]

While government agencies strive to answer all FOIA requests in an efficient and timely manner, agencies do not have to engage in further effort by researching or analyzing beyond what is necessary to respond to a specific request.[60] In addition, while FOIA allows an

---

[54] DOJ GUIDE TO FOIA, *supra* note 2, at 1; *See generally* U.S. GENERAL SERVICES ADMINISTRATION OFFICE OF CITIZEN SERVICES AND COMMUNICATIONS, YOUR RIGHT TO FEDERAL RECORDS (2004) [hereinafter GSA]); *See also id.* at 3–4 (explaining how to file a request).

[55] DOJ GUIDE TO FOIA, *supra* note 2, at 6.

[56] H.R. REP. NO. 104-175, at 19 (1995).

[57] 5 U.S.C. § 552(b)(1)–(9); 5 U.S.C. § 552(c)(1)–(3); *see also* GSA, *supra* note 54, at 1.

[58] GSA, *supra* note 54, at 1.

[59] GSA, *supra* note 54, at 1–2. To view a sample FOIA request letter, see *id.* at 4.

[60] *See id.* at 2 ("FOIA does not require [agencies] to do research for [the requester], analyze data, answer written questions, or in any other way create records in order to respond to a request.").

individual to request records in a specific form or format,[61] FOIA does not address whether a government agency must disclose a file's metadata in addition to the corresponding records requested.[62] However, to mandate the production of metadata by requiring agencies to provide records in native format would be unreasonable and impractical. It is therefore necessary to find the appropriate balance between the public's interest in obtaining access to government records and the necessity of safeguarding sensitive information. To do so is no easy task, especially since the text of the Act is silent on the issue of metadata production.[63] However, to determine whether metadata should be produced under FOIA, courts and agencies should follow standard methods of statutory interpretation and be guided solely by the language of FOIA in order to determine the meaning of what Congress enacted.[64]

B. *The Statutory Language of FOIA Provides the Procedure to Govern Records Requests*

Any dispute as to whether metadata should be produced under FOIA should be governed solely by FOIA and not subject to the Federal Rules of Civil Procedure.[65] It is well accepted that when

---

[61] *See* 5 U.S.C. § 552(a)(3)(B) (emphasis added) ("[A]n agency shall provide the record in *any form or format requested* by the person.").

[62] Although metadata may be considered a public record, there should not be a presumption that metadata should be reproduced every time that it is requested under FOIA. To do so would not only be impractical but it would create the potential for certain personally identifiable information and national security information to be released and certain federal policies to be publicly misconstrued. This could lead not only to negative consequences for citizens at an individual level, but could also have severe national security implications—such as unintentionally providing adversaries of the United States with highly sensitive defense information.

[63] The only portion of the Act that may allude to metadata is in 5 U.S.C. § 552(a)(3)(B), which allows an individual to request a record "in any form or format." However, metadata is neither mentioned in the Act nor in the comments to the Act.

[64] *See* United States v. Am. Trucking Ass'ns, 310 U.S. 534, 542 (1940) ("In the interpretation of statutes, the function of the courts is easily stated. It is to construe the language so as to give effect to the intent of Congress.").

[65] While the FRCP do not explicitly mention metadata except in one advisory comment to the rules, courts have held that metadata is subject to the general rules of discovery. *See* FED. R. CIV. P. 26, 2006 Advisory Committee Note. Therefore, in order to comply with the general rules of discovery, the FRCP, the *Sedona Principles*, and case law have all stressed the need for parties to confer in discovery proceedings to determine whether either party

legislation is brought before a court, the court should interpret that legislation by looking to the statutory text and only to the legislative history if it needs to resolve textual ambiguity.[66] Following this maxim, courts should look only to the text of FOIA to understand the procedures an agency must follow in order to comply with a FOIA request and not interpret into the statute other rules and regulations that Congress did not intend.

The only reference to the FRCP within the text of FOIA is in exemption (b)(5).[67] This reference to ordinary litigation shows that Congress knew how to make references to civil litigation within the Act and where it did not do so, the FRCP should not govern. If Congress thought it was obvious that the other provisions of FOIA

---

desires metadata. *See id.*; THE SEDONA PRINCIPLES, *supra* note 20, at 60 cmt. 3.d; Aguilar v. Immigration & Customs Enforcement Div. of U.S. Dep't. of Homeland Sec., 255 F.R.D. 350, 355 (S.D.N.Y. 2008). Because of the complexities and increase in use of electronically stored information, FRCP 26(f) was "amended to direct the parties to discuss discovery of electronically stored information during their discovery-planning conference." FED. R. CIV. P. 26, 2006 Advisory Committee Note. The conference is to be held in order to analyze the facts and circumstances of a given case to determine what, if any, type of electronically stored information would be beneficial and admissible in trial. *Id.*

It is argued that the decision to produce metadata should be a party-oriented process where parties discuss the topic at the outset of litigation in order to avoid the expense and delay of searches or productions using inappropriate or unusable forms. *See Aguilar*, 255 F.R.D. at 358 ("Rule 26(f) requires that the parties meet to confer to develop a discovery plan. That discovery plan must discuss 'any issues about disclosure or discovery of ESI, *including the form or forms in which it should be produced*. In fact, the commentary to the rule specifically notes that whether metadata should be produced may be among the topics discussed.'").

If a form of production is not specified by a party prior to court involvement, the producing party must produce the ESI in the form in which it is "ordinarily maintained" or "in a reasonably usable form, taking into account the need to produce reasonably accessible metadata that will enable the receiving party to have the same ability to access, search, and display the information as the producing party where appropriate or necessary in light of the nature of the information and the needs of the case." THE SEDONA PRINCIPLES , *supra* note 20, at ii; FED. R. CIV. P. 34(b). While courts have not conclusively decided whether metadata needs to be produced in all electronic document production, it is now generally accepted that metadata is considered to be integral to an electronic record. *See* Breen, *supra* note 35, at 439 ("The current trend is to require production of electronic information in native format with metadata intact.").

[66] *American Trucking Ass'ns*, 310 U.S. at 543.

[67] The exemption states that FOIA does not apply to matters that are "inter-agency or intra-agency memorandums or letters which would not be available to a party other than an agency *in litigation* with the agency." 5 U.S.C. § 552(b)(5) (emphasis added).

were to be governed by the same principles as ordinary litigation, there would have been no reason to include the reference in (b)(5).[68]

If the FRCP were to govern requests under FOIA, a government agency would be compelled to produce certain metadata every time it was requested under FOIA, regardless of whether the metadata was readily reproducible or of genuine public interest.[69] While an agency should not produce requested documents in a completely unusable form, it should not automatically have to produce all metadata associated with such documents.[70] Though Congress created the Act to

---

[68] In addition, if FOIA were to be governed by the FRCP, then multiple provisions within FOIA would either conflict with or be redundant to the FRCP. The following are examples of provisions within FOIA and provisions within the FRCP that conflict or if read together, would make one or the other superfluous:

- *Answering a Complaint*: 5 U.S.C. § 552(a)(4)(C) requires that an agency must answer a complaint within thirty days of notice unless otherwise good cause is shown. However, FED. R. CIV. P. 12(a)(1)(c) requires a party to reply within twenty-one days from the date of service, unless an order specifies a different time.

- *Contents of the Request*: 5 U.S.C. § 552(a)(3)(B) requires an agency to provide the requested documents in any form or format that is requested if it is readily reproducible. Somewhat similarly, FED. R. CIV. P. 34(b)(1)(C) allows for the requesting party to specify the form or form in which it is to be produced.

- *Response to a Request*: 5 U.S.C. § 552(a)(6)(A)(i) requires for a party to respond within twenty days of receipt to a request. However, FED. R. CIV. P. 34(b)(2)(A) requires the producing party to respond within thirty days of receipt to a request.

- *Attorney's Fees*: 5 U.S.C. § 552(a)(4)(E) allows for the court presiding over the case to assess attorneys fees to the United States if the requesting party has substantially prevailed on the claim. However, FED. R. CIV. P. 54(d)(2)(A) allows for attorneys fees to be considered only if there was a motion made.

[69] *See* FED. R. CIV. P. 34, 2006 Advisory Committee Note (explaining that in a motion to compel "the court is not limited to the forms initially chosen by the requesting party, stated by the responding party, or specified in this rule for situations in which there is no court order or party agreement").

[70] Rule 34(b) allows the requesting party to "specify the form or forms in which electronically stored information is to be produced" if it is relevant to the claim or defense of any party and is not privileged. *See* FED. R. CIV. P. 34(b)(1)(C). A typical request may

promote government transparency, it also recognized that some information, if disclosed, would violate an individual's constitutional right to privacy or undermine certain policies regarding national security.[71] Therefore, Congress wrote exemptions into the Act that allow an agency to withhold certain information as well as an exception to production if the information requested is not readily reproducible.[72] Using the FRCP to govern the process of record production under FOIA could lead to an inaccurate conclusion of what should and should not be reproduced. Because of this inaccuracy, a court could incorrectly compel the production of metadata that

---

look like a request to produce documents in TIFF format with a load file containing relevant metadata. *Aguilar*, 255 F.R.D. at 355. When a party produces a collection of static images, typically as a TIFF or PDF file, the images are usually not searchable or readily usable. *Id.* As a result, it is useful to provide accompanying metadata with the collection of static images so that they can be searchable and usable by the requesting party.

However, parties have a few options if they decide to produce images in static image form. The producing party can (i) avoid making load files by producing documents in their native format, which will include metadata, or (ii) produce the collection of static images in TIFF or PDF format with accompanying load files, which may or may not include metadata. In addition, courts have found that when an application is more interactive, the metadata becomes more important to understanding the application's output. *See id.* See also *infra* Parts IV & V, for consequences and recommended forms of production in the FOIA context.

Per Rule 34(b)(2)(D), the producing party then has the option to either produce or object to the format requested. FED. R. CIV. P. 34(b)(2)(D). However, if the producing party objects, it must "state the form or forms it intends to use" in producing the requested ESI. *Id.* If the requesting party disagrees with the mode of production that the producing party has stated it is going to use, the requesting party must first attempt to confer and resolve the conflict. *See* FED. R. CIV. P. 37. If the parties are unable to come to an agreement, the requesting party may then make a motion to compel production. The court will then balance the probative value of the proposed discovery and its potential burden on the producing party. FED. R. CIV. P. 26(b)(2)(C).

If the producing party can make a showing that the ESI requested would not be easily accessible and would thus pose an undue burden on the producing party, it may not be required to provide such discovery. FED. R. CIV. P. 26(b). However, *a court can still compel production despite the undue burden* if it decides that it is for good cause and relevant to the matter involved. *See Aguilar*, 255 F.R.D. at 354. When this occurs, the court may specify the conditions of how the discovery must be produced. FED. R. CIV. P. 26(b).

[71] S. REP. NO. 104-272, at 30 (1995). *See also infra* note 144 and accompanying text.

[72] *See generally* 5 U.S.C. § 552(b).

includes personally identifiable information (PII)[73] and/or other sensitive information that should not otherwise be disclosed to the public.[74] Therefore, when determining whether metadata should have to be produced under FOIA, a court should look solely to the text of the Freedom of Information Act.

C. *"Any Form or Format" Necessarily Excludes Native Format*

Despite FOIA stating that a record should be produced in "any form or format requested if the record is readily reproducible,"[75] agencies should not be required to produce records in native format—to do so would be impractical and unreasonable. Nevertheless, one scholar has opined that there should be a presumption that records be produced in native format with accompanying metadata for all requests.[76] This was developed in response to the increasing number of state decisions that have spoken to the issue of metadata production under state public records laws.[77] While the model may be appropriate when applied to state public records laws, it should not be interpreted as a realistic approach to FOIA at the federal level.[78] While the model is properly based on the purpose behind public records laws, [79] requiring the production of electronic documents with

---

[73] *See* Andrew Hotaling, *Protecting Personally Identifiable Information on the Internet: Notice and Consent in the Age of Behavioral Targeting*, 16 COMMLAW CONSPECTUS 529, n.13 at 530 (2008).

[74] *See infra* Part III.

[75] 5 U.S.C. § 552(a)(3)(B).

[76] *See* Kozinets, *supra* note 12, at 23 (finding that the "native version of an electronic public record, including metadata, should be furnished by public entities upon request").

[77] *See id.* at 25–28.

[78] It should be noted that in his article, Mr. Kozinets does provide an exception to his general model in that a government agency may withhold information if it can show, "supported by specific facts, that release of the electronic record will cause substantial harm to recognized interests of privacy, confidentiality, or other compelling government interests." *Id.* at 23. However, just releasing a document in native format will create such a substantial risk of harm to recognized interests of privacy, confidentiality, and other compelling government interests. *See infra* Part IV.

[79] The purpose behind public records laws is to allow the public the ability to monitor their officials, help to ensure public legitimacy and confidence, and promote participation by the public and democratic principles. *See* Kozinets, *supra* note 12, at 29; *see also* DOJ GUIDE

accompanying metadata in native format increases the potential for individuals to un-redact and abuse information regarding the personally identifiable information of individual citizens and sensitive national security information.

IV. CONSEQUENCES OF PRODUCING RECORDS IN NATIVE FORMAT

Current redaction techniques used for paper records are not suitable for redacting sensitive and confidential information from electronic records. It could be argued that any exempt information in a record can just be redacted from the native format of the document prior to its release[80] and that redacting the information from the native file would allow the party receiving the records the ability to see the record with accompanying metadata without seeing information that may be exempt under other parts of the Act.[81] This argument comes from the fact that when records were printed in paper form before the existence of metadata, the paper document could be redacted by producing a black bar or Xs over the exempt information.[82] While this process of redaction worked during the days of pen and paper, the complexities of electronic information and computer forensics have rendered this method of redaction ineffective for electronically stored documents. [83] Any information that is redacted from the native file of a record is never fully deleted from the

---

TO FOIA, *supra* note 2, at 1 (citing Presidential Memorandum for Heads of Executive Departments and Agencies Concerning the Freedom of Information Act, 74 Fed. Reg. 4683 (Jan. 21, 2009)).

[80] *See* Kozinets, *supra* note 12, at 26 ("If the native files contain confidential, private, or other protected or privileged metadata, the agency can withhold information that is otherwise exempt from disclosure."); *but cf.* U.S. District Court of Montana, *How to Successfully Redact Document Text*, MONT. LAWYER, Sept. 2008, at 26 ("To redact text, its metadata – literally the information about the data – must be removed.").

[81] *See generally* 5 U.S.C. § 552(b).

[82] *See* MASTERS, *supra* note 7, at 213–14 ("With paper documents, sensitive information was redacted with overlay tapes, or more commonly by using a felt-tip marker . . . . Keep in mind that the redacted information [from an electronic record] might be recoverable.").

[83] *See* Oliver Fuchsberger, *IT Tips for Ediscovery Best Practices*, WYO. LAWYER, August 2007, at 32, 34, *available at* http://wyomingbar.org/pdf/barjournal/barjournal/articles/IT_Tips.pdf ("It is not possible to redact or number native files."); Jeremy R.T. Lewis, *Electronic Access to Public Records*, *in* HANDBOOK OF PUBLIC INFORMATION SYSTEMS, 197, 207 (G. David Garson ed., 2000).

file and can be recovered.[84] As a result, this section first explores how current redaction techniques are not suitable for documents produced in native format and then explores what consequences would result if records were required to be produced in native format.

A. *Redaction Techniques Are Not Effective for Native Format*

Requiring that a government agency produce records in native format opens the floodgates for the inadvertent release of confidential information to any individual, including an adversary, by a government agency.[85] Many documents and records circulated throughout federal agencies are created and filed electronically.[86] Software systems, such as Adobe Acrobat, have developed applications that allow for the publishing and producing of documents in a standard format that consists of a scanned image of the document.[87] In addition, multiple government agencies use Redax 3.0,[88] which allows "the fullest possible disclosure to the public . . .

---

[84] MASTERS, *supra* note 7, at 213–14. *See also* ACCESS TO GOVERNMENT IN THE COMPUTER AGE: AN EXAMINATION OF STATE PUBLIC RECORDS LAWS, *supra* note 45, at 25, noting the differences in paper and electronic records. Burnett finds that "[m]ost electronic materials do not exist in paper form. In fact, one of the primary reasons for having electronic records is to conveniently store more information than one could reasonably store in paper form." In addition, Burnett continues noting the differences between paper and electronic file destruction, finding that "while paper records can be thrown away, shredded, burned, or otherwise completely destroyed, e-records can be much harder to get rid of . . . . Even when [a record] is overwritten . . . the data may still exist." *Id.* at 25–26. While an electronic and paper version of a record may look exactly the same, "an e-record and a paper record are drastically different." *Id.* at 26; Kaplan, *supra* note 7 ("It used to be that to redact a document, you took a black magic marker or sticky white tape and . . . problem solved. Today, however, redaction is a critical feature of document management, especially given heightened national security and personal privacy concerns post-Sept. 11."); Alan Blakley, *Differences and Similarities in Civil Discovery of Electronic and Paper Information*, FED. LAWYER, July 2002, at 32, 32 ("[The Southern District of New York's] adaptation of paper discovery principles to electronic discovery demonstrates a grasp of the issues encountered.").

[85] *See infra* Part IV.B.

[86] KAPLAN, *supra* note 7.

[87] *Id.* It is estimated that "approximately 200 to 300 government agencies, including courts, have adopted PDF as a standard electronic format."; *see, e.g.*, MASTERS, *supra* note 7, at 213; *see also* NSA, *supra* note 7.

[88] For more information regarding how Redax 3.0 works, *see Redax Enterprise Server*, APPLIGENT DOCUMENT SOLUTIONS (Mar. 5, 2012), http://docs.appligent.com/docs-res.

[while ensuring that the] documents [are] in compliance with both the Privacy Act and the Freedom of Information Act."[89] The process of converting a file to a PDF ensures a high level of security[90] because metadata is typically removed during the conversion.[91] The most secure way to ensure that any exempt information is kept confidential through the redaction process is to convert the file to a PDF.[92]

In electronic files, information that is covered or obscured is never really removed from the file.[93] Replacing exempt information with Xs, creating a black bar over the text, or changing the font color to white has been thought to be an effective way to eliminate the inadvertent transmission of exempt information.[94] However, because of recent advances in computer forensics and technology, these methods are

---

[89] *See PR: U.S. Department of Justice Selects Appligent Redax for PDF Redaction*, PLANET PDF, (Mar. 5, 2012), http://www.planetpdf.com/mainpage.asp?webpageid=2450 (explaining that Redax is used by a number of government agencies, including "the Secret Service, the Internal Revenue Service, the Food and Drug Administration, the National Reconnaissance Office and the Department of Energy"). The software allows these "agencies to implement a cost-effective redaction solution that is easily incorporated into their existing workflows and document management processes." *Id.* (internal quotations omitted). In addition, "Redax allows a simple box to be drawn over the irrelevant [or exempted] information and a code to be inserted, referring to the specific court rule that justifies the redaction." KAPLAN, *supra* note 7.

[90] KRAUSE, *supra* note 28.

[91] *See id.* ("Turning a file into a PDF or using a metadata-stripping tool strips out . . . metadata including data associated with a 'track changes' feature or comments."); *see also* KAPLAN, *supra* note 7.

[92] *See infra* Part V.

[93] *See* U.S. DISTRICT COURT OF MONTANA, *supra* note 80, at 26, finding that "[i]n an electronic file, the obscured text still lurks beneath the highlighter box and can be readily recalled. The text is hidden, not excised."

[94] *See id.* ("[S]imply pasting the [redacted] text into a Word file [can reveal] the hidden information. Changing the text color to white so it disappears against the white screen/paper is similarly ineffective. To redact text, its metadata . . . must be removed."); *see also* Jason Hart & Walter R. Houser, *Software Helps Tidy up FOIA Responses*, GOVERNMENT COMPUTER NEWS (Oct. 27, 1997), *available at* http://gcn.com/Articles/1997/10/27/Software-helps-tidy-up-FOIA-responses.aspx?p=1 (last visited Mar. 5, 2012) ("A quick save in Microsoft word makes an addendum to the file without obliterating the deleted information. The Microsoft Windows Notepad still holds deleted text for an enterprising snooper to find. Other word processors behave in similar ways . . . ."); *see further* BLAKLEY, *supra* note 83, at 34 ("System data and metadata may contain material that cannot be redacted, and some 'deleted' data themselves may survive attempts to redact.").

not as effective as one may assume.[95] As a result, some have urged that documents should be sent as hard copies so that the metadata, which "may reveal privileged or confidential information in various ways . . . cannot be transmitted."[96]

The concern over the possibility of revealing redacted text has grown substantially in ordinary litigation as the inadvertent disclosure of privileged information has become increasingly common.[97] In ordinary litigation for example, there are significant penalties for lawyers who seek to uncover redacted information or who fail to disclose their receipt of the privileged information.[98] Regardless of these penalties, even if certain information were disclosed during the discovery process, it would probably not have much of an effect on the outcome of the litigation.[99] The same cannot be said for the disclosure of certain information to an individual who has the propensity to abuse the information to the detriment of individual citizens and the nation as a whole.

If records requested under FOIA were required to be produced in native format, any confidential information that was redacted by the agency has the possibility of being recovered by an enterprising adversary. Even without the appropriate redacting software, redacted information can be revealed from a file in PDF format.[100] However,

---

[95] These methods simply cover up the underlying metadata as opposed to removing it from the document. *See generally*, U.S. DISTRICT COURT OF MONTANA, *supra* note 80; *see also infra* note 100 and accompanying text.

[96] BENNETT & CLOUD, *supra* note 17, at 474–75.

[97] Lisa C. Wood & Marco J. Quina, *The Perils of Electronic Filing and Transmission of Documents*, 22 ANTITRUST 91 (2008). The American Bar Association (ABA) has even proscribed rules for handling the inadvertent transmission of metadata in the Model Code of Professional Conduct. *See* MODEL RULES OF PROF'L CONDUCT R. 1.6 (2009); *see also* BENNETT & CLOUD, *supra* note 17, at 473.

[98] BENNETT & CLOUD, *supra* note 17, at 477; Rule 8.4(c) states, "It is professional misconduct for a lawyer to . . . engage in conduct involving dishonesty, fraud, deceit or misrepresentation." MODEL RULES OF PROF'L CONDUCT R. 8.4(c) (2009).

[99] *See, e.g.*, Cont'l Grp. Inc. v. KW Prop. Mgmt., LLC, 622 F. Supp. 2d 1357, 1373 (S.D. Fla. 2009) *order clarified*, 09-60202-CV-COHN, 2009 WL 3644475 (S.D. Fla. Oct. 30, 2009) (finding that the defendant's destruction of all metadata evidence embedded in the documents did not amount to bad faith because "metadata evidence is not obvious to non-computer professionals").

[100] *See, e.g.*, Wood & Quina, *supra* note 97, at 91 (explaining how the FTC "inadvertently made public . . . sensitive business information, including marketing and negotiation strategies and plans to close an number of stores"); *see further id.* (citing Dan Shea,

given the problems arising out of inadvertent disclosures, most agencies now use redaction software, such as Redax 3.0 or Adobe Acrobat 8, to prevent the future inadvertent disclosure of information.[101] While not 100% secure, these software applications provide an effective way to ensure that redacted information cannot be recovered.[102] In doing so, however, these applications work to remove metadata and confidential information from static image files, not native files.[103] If individuals could recover redacted information from PDF files, then it is without a doubt possible–and much easier–to recover information redacted in native files. Even more disturbing is the possibility of confidential information being released to adversaries simply because government agencies would be required to produce documents in native format so that metadata (while sometimes wholly irrelevant) can be given to the requester.

B. *The Potential Consequences Arising from Recovering Redacted Information from Native Files*

It could be suggested that providing an exception—which would allow a government agency to withhold disclosure if it determines the disclosure would compromise personally identifiable information or national security information—to a general rule of mandating the

---

Military Gaffe Results in Classified Data Leak, PLANET PDF, (May 6, 2005), http://www.planetpdf.com/enterprise/article.asp?ContentID=Militarygafferesultsinclassifieddataleak&gid=7049, and explaining that while the Pentagon attempted to redact some information by visually impairing the confidential information, "the Pentagon inadvertently revealed the identities of individuals and other confidential information by posting a report of an investigation into the killing of an Italian intelligence agent in Baghdad").

In a case against the NSA for wiretapping, the redacted text of a legal brief could be recovered by "cop[ying] and past[ing] inside some PDF readers." Declan McCullagh, *AT&T Leaks Sensitive Info in NSA Suit*, CNET NEWS.COM, (May 26, 2006), http://www.news.com/2100-1028_3-6077353.html. In 2006, certain portions of a legal brief were "electronically blacked out to protect . . . sensitive material . . . [b]ut the passages [could] be viewed by simply pasting the document into a word processing program." Adam Liptak, *Prosecutors Can't Keep a Secret in Steroid Case*, N.Y. TIMES, (June 23, 2006), *available at* http://www.nytimes.com/2006/06/23/us/23leak.html; *see also* WESCOTT, *supra* note 15, at 5.

[101] For an explanation of the various software programs used by agencies, see HART & HOUSER, *supra* note 94.

[102] *Id.*

[103] *See* MASTERS, *supra* note 7.

production of records in native format with metadata would not be practical.[104] Every request, if provided in native format, would risk the possibility of substantial harm—for the government to have to provide such a showing for every request under FOIA would not only be burdensome, but highly impractical.[105] In addition, the following hypothetical situations demonstrate how the release of government records in native format would allow an adversary to recover and abuse confidential information related to an individual citizen's or the nation's security.

Suppose, for example, someone requests records from the Department of Defense (DOD) in electronic format with accompanying metadata that relate to certain military personnel stationed in the Middle East. Suppose also that the individual who requests the records is an affiliate of a terrorist organization that opposes United States military operations in the Middle East.[106] Despite certain exemptions within FOIA that preclude the release of national defense information, the requester has requested a multitude of documents which, individually, do not reveal information that would fall within any exemption.[107] Some of the documents, however, contain small portions of text, which would have to be redacted for exemption reasons, but because most of the records can be produced

---

[104] KOZINETS, *supra* note 12, at 23 ("[A] showing that is supported by specific facts, that release of the electronic record will cause substantial harm to recognized interests of privacy, confidentiality, or other compelling government interests.").

[105] FOIA only mandates that records be reproduced if they are "readily reproducible." *See* 5 § U.S.C. § 552(a)(3)(B) (2006). In addition, given the state of backlogs experienced by multiple government agencies, to increase the workload by mandating a showing of substantial harm with every document is highly impractical. For an example of an explanation and summary of government agency FOIA backlogs, see Gary M. Stern, *Chief FOIA Officer's Report*, NATIONAL ARCHIVES AND RECORDS ADMINISTRATION 1, 9 (Jan. 31, 2011), http://www.archives.gov/foia/reports/chief-foia-officer-report-2011.pdf; *see also* Debra Wall, *Report Show Federal Agency FOIA Backlogs Growing*, THE NATIONAL COALITION FOR HISTORY (Mar. 5, 2012), http://historycoalition.org/2011/07/05/report-show-federal-agency-foia-backlogs-growing.

[106] While it is presumed that the government has intelligence regarding adversaries to the United States, it is possible that one can go undetected until committing or attempting to commit a terrorist plot.

[107] *See* 5 U.S.C. § 552(b)(1)(A) (2006). The concept of the mosaic theory is beyond this Article. *But see* David E. Pozen, *The Mosaic Theory, National Security, and the Freedom of Information Act*, 115 Yale L.J. 628 (2005); *see also* PATRICE MCDERMOTT, WHO NEEDS TO KNOW? THE STATE OF PUBLIC ACCESS TO FEDERAL GOVERNMENT INFORMATION 137 (2008).

without issue, the DOD cannot simply refuse to comply with the request.[108] If the DOD was required to produce the records in native format, the adversary could simply recover the redacted information and obtain information regarding the military personnel (including any comments written on the track changes functions of the record) within each record. By combining the information from all of the various records requested,[109] the government could simply hand over the information needed by an adversary to cause great harm to the United States.[110]

Let us take another example. Suppose someone requests documents from Customs and Border Protection (CBP) regarding the status of a newly implemented border control program that would establish more security at the border to protect against the growing issue of the drug trade. This individual, while not a terrorist, is connected to a large drug cartel located outside of the United States. The information requested under FOIA is to be produced electronically with accompanying metadata. As such, CBP properly

---

[108] *See* 5 U.S.C. § 552(a) (2006), *amended by* 5 U.S.C. § 552(a)(4)–(a)(8) (Supp. V 2011).

[109] See GIDIERE, *supra* note 3, at 176–77, commenting on the ability of pulling together separate pieces of information in order to create a larger picture. He states, "[o]n the one hand, this ability to assemble information can increase the efficiency and integrity of government programs and services. On the other hand, the potential for abuse and unintended consequences is evident." *Id.* See also CHUMBLER, *supra* note 45, at 73, further commenting on one's ability to assemble smaller pieces of data into a larger collection. She argues that information from one source can be combined with another and put into different contexts that are "far removed from their original purposes." *Id.* In addition, Chumbler argues, "information received from one agency–even when redacted to remove personal information–can potentially be combined with other publicly available data to arrive at the very information that was redacted." *Id.*

[110] See GIDIERE, *supra* note 3, at 353, emphasizing that FOIA does not permit selective disclosure "so even a suspected terrorist oversees arguably has the same rights to request and receive federal information as an American citizen . . . . Moreover, frequently requested records must be made available in agency electronic reading rooms, thereby making access even more convenient to a potential terrorist." *Id.* For these reasons, Gidiere states, "the use of exemptions is essential to preventing widespread dissemination of homeland security information." *Id.* However, when producing records in native format, the ability to recover redacted information that would fall under the exemptions would essentially make such exemptions rather superfluous and still provide some information to potential terrorists. *Id.* It must also be noted that confidential records that wholly consist of national security information would not be released due to the exclusions within FOIA. *See* 5 U.S.C. § 552(b)(1)(A) (2006). However, what this example explores is the possibility of multiple documents individually containing fairly little information regarding national security that may, collectively, reveal much more than any government agency would like for an individual requester to see. *See* GIDIERE, *supra* note 3, at 353.

responds to FOIA request and produces the records in native format in order to comply with the request. CBP properly redacts information regarding national security concerns at the border, such as certain locations that are less secure. The requester receiving this information has the ability to un-redact this information and feed it to their counterpart in a foreign country, thereby increasing the amount of illegal contraband let into the country.

Let us take as our last example that of an individual who requests documents about a recent hospital initiative developed by the Department of Health and Human Services (HHS). The initiative was developed to aid those who had been admitted to a hospital with a certain illness in the past year by granting them $500 each. Unfortunately, you contracted this illness last year. Suppose that the individual requested all FOIA records pertaining to this initiative and their accompanying metadata. Because some of the records contained patients' information including names, addresses, and social security numbers, any part of the record that had that information was redacted by the agency in order to comply with exemption (b)(6) of FOIA.[111] However, if the records were to be produced in native format, the requester could simply un-redact that information and suddenly have access to your name, address, and social security information.

These hypotheticals, while not extremely fact-specific, are simply meant to provide possible examples of the negative consequences that could result from producing records in native format and the ability of an adversary to recover redacted information. Given the abundance of records containing sensitive national security information and the amount of information that government agencies have on each individual throughout the United States, any person making a request under FOIA can recover personal and confidential information about military operations or you, the reader, simply by requesting documents under FOIA, which would require that metadata must be produced in native format.[112]

---

[111] *See* 5 U.S.C. § 552(b)(6) (2006) (providing that an agency does not have to reproduce records that consist of "personnel and medical files and similar files the disclosure of which would constitute a clearly unwarranted invasion of personal privacy").

[112] *See generally* POZEN, *supra* note 106 (discussing the possibility of linking a bunch of information about an individual together); *see also* CHUMBLER, *supra* note 45, at 72 ("The transformation of public records from paper documents to electronic media has dramatic implications for personal privacy."). Citing the Florida Supreme Court, Chumbler noted, "digital storage and transfer of information changes how information can be manipulated and retrieved." *Id.* Previously obscure information can be located quickly and anonymously for essentially no cost . . . . These and other issues raise deep concerns about the use of

C. *Contextual Inaccuracies and the Potential for Misconstrued Public Opinion*

In addition to the possibility of an adversary receiving information that was redacted but easily recovered because the record was produced in native format, there is always the possibility of metadata being produced that, while seemingly not that important, can be contextually inaccurate and misconstrued by the public.[113] There is a rising concern in ordinary litigation over the authenticity of information, which, if inaccurate, could lead to possible evidentiary problems.[114] One can argue that "metadata is the only source of evidence that bears on authenticity," yet metadata itself can even be altered.[115] As noted at the Sedona Conference,[116] "the ease of transmitting electronic data and the routine modification and multi-user editing process may obscure the origin, completeness, or accuracy of a document."[117]

For example, when an individual creates a document in a word processing software, somewhere embedded in the file is the author of the document. The author, however, is attributed to the name on the computer, typically the owner.[118] If an individual were to use someone

---

information for purposes other than those for which the information was initially provided." *Id.* at 72.

[113] For example, the time an e-mail was sent or who authored a document. *See generally* DOJ GUIDE TO FOIA, *supra* note 2, at 366; *see also* AFGE v. U.S. Dep't of Health & Human Servs., 63 F. Supp. 2d 104, 108 (D. Mass. 1999) (holding that the release of draft versions of documents "could cause harm by providing the public with erroneous information"), *aff'd*, AFGE, Local 1146 v. U.S. Dep't of Health & Human Servs., No. 99-2208, 2000 U.S. App. LEXIS 10993, at *3 (1st Cir. Mass. May 18, 2000).

[114] *See* PAUL & NEARON, *supra* note 6, at 106.

[115] *Id.*

[116] The Sedona Conference Working Group on Best Practices for Electronic Document Retention and Production is "designed to bring together some of the nation's finest lawyers, consultants, academics, and jurists to address current problems in the areas . . . in need of a 'boost' to advance law and policy." THE SEDONA PRINCIPLES, *supra* note 20, at i.

[117] *Id.* at 5.

[118] WESCOTT, *supra* note 15, at 9.

else's computer to create a document, the author as shown through the metadata would be the owner of the computer, despite the owner possibly having no knowledge that the document was ever created.[119] Producing documents in native form with metadata can lead to severely misconstrued public opinion of a certain policy developed by an agency. In addition, if a document is produced in native format, the individual receiving the record can simply alter the metadata herself thereby creating falsified comments, changing the author of the document, or changing the time a document was sent.[120] This, too, has the potential for severe consequences.

For sake of clarification, the following examples present the possible consequences that may result from the release of native format documents. First, suppose a record was created on June 1 but was sent via e-mail to another individual to store in records on June 10. The document reproduced in response to the request was the one from the latter individual's computer, which re-stamped the document with the new date.[121] If the requester sought the documents to show that an agency failed to implement a policy on June 5, the metadata showing the date would be inaccurate and could have detrimental effects to the agency's reputation.

Now suppose that a government official, now head of the entire agency, created a document in 2005 as a template for a document that is to be used to record certain incidents resulting from a natural disaster. Suppose after a huge natural disaster an individual inaccurately describes the event and such inaccuracies create huge problems with the response team, resulting in more death and devastation to the victims. The author of the document, as shown by the metadata, may still be the head of the department who created the template in 2005 despite the lower official being the individual who so grossly misreported the data causing the controversy. If the metadata

---

[119] *See* Craig Ball, *Beyond Data About Data: The Litigator's Guide to Metadata*, STO01 ALI-ABA 781, 811 (2011) ("Computers may be shared or unsecured and passwords lost or stolen. Software permits alteration of documents.").

[120] *See* BENNETT & CLOUD, *supra* note 17, at 487 ("Native file documents may be more easily altered or manipulated by users."); FUCHSBERGER, *supra* note 82, at 34 ("[N]ative files can . . . be accidentally or intentionally modified.").

[121] It is possible that an individual who requests records under FOIA seeks information regarding the time certain communications were shared between individuals; however, "now documents may generate a new date each time they are opened[,]" possibly confusing the requester into believing that the possible alteration of a document could be the creation date. *See* BALL, *supra* note 118, at 811.

were given, it would give a false impression of who created the document and could result in devastating consequences, both professionally and personally, to the head of the department.[122]

These limited examples provide potential issues that can arise out of the contextual inaccuracies of producing metadata in native format. What may seemingly be a minor detail in a document can have severe consequences if one person can manipulate a document to create an opinion by a government official or otherwise to create a backlash by the public to a program implemented by any agency.[123]

V. A PROPOSED MODEL: WHEN AND HOW METADATA SHOULD BE PRODUCED

While some metadata may not need to be produced, a government agency should try to respond to a FOIA request in a way that best responds to what the requesting party specifies.[124] In addition, when discussing FOIA, it "is important to . . . note that the President and Attorney General have issued memoranda to all agencies emphasizing that FOIA reflects a 'profound national commitment to ensuring an open Government' and directing agencies to 'adopt a presumption in

---

[122] *See* ROBERTSON, *supra* note 16, at 204; *see also* WESCOTT, *supra* note 15, at 15 ("For example, when a new employee uses a word processing program to create a memorandum by using a memorandum template created by a former employee, the metadata for the new memorandum may incorrectly identify the former employee as the author."). *See also* THE SEDONA PRINCIPLES, *supra* note 20, at 5, providing that:

> there is growing use of collaborative software that allows for group editing of electronic data, making authorship determination more difficult. Finally, while electronically stored information may be stored on a single location, such as a local hard drive, it is likely that such documents may also be found on high-capacity, undifferentiated backup tapes, or on network servers—not under the custodianship of an individual who may have "created" the document.

See further Ky. Speedway, LLC v. Nascar, Inc., No. 05-138-WOB, 2006 U.S. Dist. LEXIS 92028, at *24 (E.D. Ky. Dec. 18, 2006) ("Depending on the format, the metadata may identify the typist but not the document's author, or even just a specific computer from which the document originated or was generated.").

[123] While it is understood that there may be criminal penalties against individuals who provide inaccurate information to the media, including defamation, providing such altered information "may cause a lasting effect on the victim's reputation." Peter Meijes Tiersma, *The Language of Defamation*, 66 TEX. L. REV. 303, 310 (1987).

[124] GSA, *supra* note 54, at 2.

favor of disclosure.'"[125] In order to reconcile these two opposing principles, this section provides a framework for determining whether an agency should produce metadata in a given case. First, the requester must specifically request what metadata is desired when she files the initial records request. Second, if the metadata is not exempt under exemption (b)(5) and is readily reproducible, it should be reproduced by the government agency. Third, if the records and metadata must be produced, they should only be produced in static image format with accompanying Bates numbers and load files.[126] This framework allows the requesting party to receive the desired records and accompanying metadata in a manner that is consistent with the language of FOIA and also protects from disclosure certain information that could have severe consequences if released to the public.[127]

A. *Metadata Must Be Specifically Requested*

If an agency is to produce metadata, the requesting party must have specifically requested the metadata[128] by including in its FOIA request what records should be accompanied by their metadata. This would enable the agency to determine what metadata would need to be produced, and would help a court review the specifics of the requested metadata in coordination with the elements below in the event that the agency fails to provide the metadata and is subject to subsequent litigation.[129] If metadata is not specifically requested, the

---

[125] DOJ GUIDE TO FOIA, *supra* note 2, at 357.

[126] A Bates number refers to the "[s]equential numbering used to track documents and images in production data sets, where each page is assigned a unique production number. Often used in conjunction with a suffix or prefix to identify the producing party, the litigation, or other relevant information." THE SEDONA CONFERENCE GLOSSARY, *supra* note 6. For instruction on how to Bates number a document, see MASTERS, *supra* note 7, at 207–12.

[127] *See supra* part IV.

[128] When filing a request, the requester should, as specifically as possible, address what metadata the requester wants and for what records the requester wants it. For an example of what an ordinary FOIA request looks like, see GSA, *supra* note 54, at 4.

[129] One who requests certain information under FOIA and whose request is denied may appeal such a denial. If the appeal is also denied, the requester can then file suit in the U.S. District Court in which they reside. *See* GSA, *supra* note 54, at 9.

agency should not have to reproduce it at all[130]—to have to go through individual records' metadata would only increase the burden on an agency and potentially worsen the backlogs that currently affect the FOIA request process.[131]

B. *When Metadata Should Not Have to Be Produced Despite Being Requested*

As this Article and the complexities of the Freedom of Information Act show, creating a general rule that records should be produced in native format to include metadata whenever requested is impracticable. In order to determine when metadata needs to be reproduced, a court should ask whether the metadata is (1) exempt and (2) whether the metadata is readily reproducible.

1. *Substantive Metadata Is Exempt under FOIA Exemption (b)(5)*

Certain metadata does not have to be reproduced under FOIA if it is pre-decisional and deliberative, thereby falling under exemption (b)(5) ("exemption 5").[132] The deliberative process privilege has been incorporated into exemption 5 of FOIA and has been invoked when an agency seeks to prevent the disclosure of draft versions of records to the public.[133] The Supreme Court has upheld the defense finding that

---

[130] *Id.* at 2–4.

[131] *See supra* note 104 and accompanying text.

[132] *See* 5 U.S.C. §§ 552(a)(3)(B), 552(b)(5) (2006), *amended by* 5 U.S.C. § 552(b) (Supp. V 2011) ("[A]n agency shall provide the record in any form or format requested by the person if the record is readily reproducible by the agency in that form or format . . . . [However,] [t]his section does not apply to matters that are . . . (b)(5) inter-agency or intra-agency memorandums or letters which would not be available by law to a party other than an agency in litigation with the agency[.]"). *See also* Margaret B. Kwoka, *The Freedom of Information Act Trial*, 61 AM. U. L. REV. 217, 234 (2011) (citing United States v. Weber Aircraft Corp., 465 U.S. 792, 799 (1984) (citing FTC v. Grolier Inc., 462 U.S. 19, 26 (1983)), (finding that the Supreme Court has explained that exemption 5 "'simply incorporates civil discovery privileges.' The three standard privileges invoked under exemption 5 are the deliberative process privilege, the attorney-client privilege, and the attorney work-product privilege. Of those, the deliberative process privilege is the most frequently used to withhold records.").

[133] The deliberative process privilege has been utilized "to encourage open, frank discussions on matters of policy between subordinates and superiors; (2) to protect against premature disclosure of proposed policies before they are actually adopted; and (3) to protect against public confusion that might result from disclosure of reasons and rationales

the privilege protects the "decision making processes of government agencies."[134]

Whether metadata falls within exemption 5 has never been addressed; however, it can be logically inferred that certain metadata will fall within exemption 5 because the "ultimate objective of exemption 5 is to safeguard the deliberative process of agencies, not the paperwork generated in the course of that process." [135] Nevertheless, in order for the defense to be utilized, the metadata must meet two criteria: (a) it must be pre-decisional and (b) it must be deliberative.[136]

a. *Substantive Metadata Can Be Considered Pre-Decisional*

If the non-electronic form of the drafting process is protected, then the same protection should be afforded to the electronic form.[137] Many courts have found that the very process of converting a draft into a final product constitutes a deliberative process and have thus held that drafts are protected from disclosure under FOIA.[138] With the

---

that were not in fact ultimately the grounds for an agency's action." DOJ GUIDE TO FOIA, *supra* note 2, at 366; *See* NLRB v. Sears, Roebuck & Co., 421 U.S. 132, 151 (1975).

[134] DOJ GUIDE TO FOIA, *supra* note 2, at 366.

[135] *See* Nat'l Wildlife Fed'n v. U.S. Forest Serv., 861 F.2d 1114, 1119 (9th Cir. 1988).

[136] DOJ GUIDE TO FOIA, *supra* note 2, at 368 ("[C]ourts have established two fundamental requirements . . . for the deliberative process privilege to be invoked. (footnote omitted). First, the communication must be pre-decisional . . . . Second, the communication must be deliberative . . ."); *see also* Mapother v. Dep't of Justice, 3 F.3d 1533, 1537 (D.C. Cir. 1993) ("The deliberative process privilege protects materials that are both pre-decisional and deliberative." (citing Petroleum Info. Corp. v. U.S. Dep't of the Interior, 976 F.2d 1429, 1434 (D.C. Cir. 1992))).

[137] The purpose behind the exemption is more relevant than the actual form in which the deliberations occurred. *See* Schell v. U.S. Dep't of Health & Human Servs., 843 F.2d 933, 940 (6th Cir. 1988) ("Because Exemption 5 is concerned with protecting the deliberative process itself, courts now focus less on the material sought and more on the effect of the material's release."). For decisions to withhold documents in paper requests, see Kidd v. Dep't of Justice, 362 F. Supp. 2d 291, 296 (D.D.C. 2005) (internal quotations omitted) (precluding the production of paper draft versions on the basis that disclosure would "inhibit drafters from freely exchanging ideas, language choices, and comments in drafting documents"); State of Mo. ex rel. Shorr v. U.S. Army Corps of Eng'rs, 147 F.3d 708, 710 (8th Cir. 1998) ("The purpose of the deliberative process privilege is to allow agencies freely to explore alternative avenues of action and to engage in internal debates without fear of public scrutiny.").

[138] DOJ GUIDE TO FOIA, *supra* note 2, at 389–90.

increased use of technology, certain software programs, like track changes in Microsoft Word,[139] now help users create substantive metadata. As noted above, substantive metadata includes metadata where users can electronically communicate on the face of a document multiple times before the record is considered final.[140] When the final document is created, all of the substantive metadata can be accepted or made invisible so that a viewer only sees the final, output display of the file.[141] The comments, revisions, and possible policy considerations that are exchanged between the two users would constitute "pre-decisional" material because they are part of the deliberative process before the final policy of an agency was decided upon.[142]

In addition, the Supreme Court has noted that even if there was no final decision that resulted from the deliberation, the substantive metadata may still be considered pre-decisional.[143] While the Supreme Court was referring to documents in print form, the same rationale should be applied to documents in electronic form. Merely using modern technologies to engage in the deliberative process electronically does not warrant the application of a different standard.

b. *Substantive Metadata Is Inherently Deliberative*

---

[139] This function is an electronic system of editing and revising that mimics the traditional, non-electronic form of drafting.

[140] *Find and Remove Metadata (Hidden Information) in Your Legal Documents*, MICROSOFT INC., http://office.microsoft.com/en-us/help/find-and-remove-metadata-hidden-information-in-your-legal-documents-HA001077646.aspx (last visited Mar. 8, 2013).

[141] *Id.*

[142] *See* DOJ GUIDE TO FOIA, *supra* note 2, at 368–72.

[143] *See* Sears, Roebuck & Co., 421 U.S. at 151 n.18, providing that:

> Our emphasis on the need to protect pre-decisional documents does not mean that the existence of the privilege turns on the ability of an agency to identify a specific decision in connection with which a memorandum is prepared. Agencies are, and properly should be, engaged in a continuing process of examining their policies; this process will generate memoranda containing recommendations which do not ripen into agency decisions; and the lower courts should be wary of interfering with this process.

See also DOJ GUIDE TO FOIA, *supra* note 2, n.80 at 370.

If there are multiple documents requested under FOIA regarding the policy decisions of a federal agency, then draft versions of those documents that include substantive metadata may be requested as well. The accompanying substantive metadata may include edits, revisions, and comments created by a different party. Deliberative is typically defined as involved in discussion[144] and considering that the very nature of certain software is to aid in fostering electronic discussion between individuals, this type of substantive metadata should fall within the definition of deliberative.[145]

2. *Metadata Should Not Have to Be Reproduced if It Is Not 'Readily Reproducible'*

If an agency determines that complying with a FOIA request would be unreasonable because the records requested are not readily reproducible with accompanying metadata due to the technical limitations of the agency, the metadata should not have to be produced.[146] FOIA states, "A court shall accord substantial weight to an affidavit of an agency concerning the agency's determination as to technical feasibility . . . and reproducibility."[147] This provision was added to the Act as part of the 1996 e-FOIA amendments and, as noted above, the purpose of these amendments was to ensure that the government's increased use of technology did not interfere with the policy of promoting government transparency.[148] However, Congress

---

[144] *See* THE NEW OXFORD AMERICAN DICTIONARY 451 (2001) (defining deliberative as, "relating to or intended for consideration or discussion").

[145] Other types of metadata, such as system metadata, are not likely to be considered deliberative and do not fall within exemption 5. In circumstances in which non-deliberative metadata is requested under FOIA, a court will have to look to the agency's ability to readily reproduce the metadata to determine whether it needs to be produced.

[146] *See* Mead Data Cent., Inc. v. U.S. Dep't. of Air Force, 566 F.2d 242, 261, n.55; *see also* FlightSafety Servs. Corp. v. Dep't of Labor, 326 F.3d 607, 613 (5th Cir. 2003) (per curiam) (finding that documents did not have to be produced because "producing it would require substantial agency resources and produce a document of little informational value"); Doherty v. Dep't of Justice, 775 F.2d 49, 53 (2d Cir. 1985) (finding that the district court was not required to analyze approximately 300 pages of documents, line-by-line); Solar Sources, Inc. v. United States, 142 F.3d 1033, 1039 (7th Cir. 1998) (finding that it would take eight work years to identify all of the documents requested).

[147] 5 U.S.C. § 552(a)(4)(B) (2006), *amended by* 5 U.S.C. § 552(a)(4) (Supp. V 2011).

[148] DOJ GUIDE TO FOIA, *supra* note 2, at 6.

included this provision within the Act knowing that some ESI would not be readily reproducible and that the government agency requested to produce such information would have the best knowledge to speak of its technical capability.[149]

Courts have referred to the ability of an agency to readily reproduce information in a certain form or format in terms of "technical capability" or "technical feasibility."[150] Whether information under a FOIA request is readily reproducible will turn on a case-by-case analysis of the fact situation provided.[151] For example, if a party requests a small number of records with accompanying metadata, then the records and metadata will likely be deemed to be readily reproducible. However, if a party requests records with accompanying metadata that amount to hundreds or potentially thousands of documents, the ability for an agency to readily reproduce the documents with metadata in a form that satisfies the party's request and still complies with the exemptions in section (b) of FOIA is likely to be called into question.[152]

FOIA only requires that an agency use "reasonable efforts" to search for and maintain documents in their original form.[153] This qualification "could relieve agencies of the obligation of releasing the original form of partially exempt records in circumstances where agencies need to handle the records in a certain form for purposes of

---

[149] S. Rep. No. 104-272, at 15 (1995).

[150] *See Nat'l Day Laborer Org. Network*, 2011 U.S. Dist. LEXIS at *8; ("'Readily reproducible' simply refers to an agency's technical capability to create the records in a particular format."); *see also* Sample v. Bureau of Prisons, 466 F.3d 1086, 1088 (D.C. Cir. 2006).

[151] For a more detailed explanation of "readily reproducible," *see* GIDIERE, *supra* note 3, at 151–54.

[152] While this Article analyzes metadata in terms of exemption 5, it must be noted that there are other exemptions within the Act that may require certain non-metadata components of documents to be redacted. *See* 5 U.S.C. § 552(b)(1)–(9) (2006), *amended by* 5 U.S.C. § 552(b) (Supp. V 2011). This could increase the difficulty of the agency's production because producing certain metadata, such as to whom an email is addressed, could be exempt under a different exemption because it contains Personally Identifiable Information (PII). This is because if a document is requested to be in native form, which would include metadata, as opposed to a static image form, then the previously redacted information may be un-redacted by the requesting party. This adds difficulty and technical complications to an agency producing records.

[153] 5 U.S.C. § 552(a)(3)(C).

redaction."[154] In addition, if accompanying metadata was mandated to be produced in every case, an agency may need to expend vast resources in order to ensure that the records reproduced not only comply with the Act, but that the metadata does as well.[155] Congress did not intend for an agency to use unreasonable efforts in electronic record production, because doing so could potentially have a significant impact on the agency's ability to perform its day-to-day functions.[156] In addition, Congress, while encouraging government transparency, did not intend for electronic production to "result in any greater expenditure of agency resources than would have occurred" in conventional paper-based FOIA requests.[157]

Although an agency may determine that it would not be able to readily reproduce documents with accompanying metadata, the information provided to the requesting party should still be produced in a searchable and usable manner. However, when a party has requested and been denied records with accompanying metadata and therefore subsequently files suit to compel disclosure, the government agency can assess its ability to readily reproduce such information and explain whether the agency is technically capable or incapable of doing so in an affidavit submitted to the court. Because the agency is in the best position to determine its technical ability, Congress instructed the court to give deference to the agency's determination by according substantial weight to the affidavit submitted. [158] Nevertheless, if the records and metadata are compelled, they should only be produced in a format that protects the integrity of the document as well as furnishes the requested information.

C. *A Requirement to Produce Records with Accompanying Metadata in Static Format Only*

---

[154] S. REP. NO. 104-272, at 15 (1995).

[155] *See supra* note 144 and accompanying text.

[156] S. REP. NO. 104-272, at 15 (1995).

[157] H.R. REP. NO. 104-175, at 22 (1995).

[158] 5 U.S.C. § 552(a)(4)(B).

While electronic records in ordinary civil litigation are typically produced in either native format or as static images,[159] given the vast differences between ordinary civil litigation and requests under FOIA, the former option should not be applied to the FOIA context.[160] If metadata is compelled by a court to be produced by an agency, the agency should only reproduce the metadata as part of the record in static form with accompanying load files.[161] To mandate that an agency reproduce a record in native format so that the metadata remains intact would be to mandate that an agency must expose itself to the risk that sensitive information is recovered by an enterprising adversary or organization.[162]

Producing a document in static form drastically decreases the possibility that any information that is redacted by an agency will be recovered.[163] This format will necessarily preclude some metadata from being reproduced.[164] However, if metadata is requested and its disclosure compelled, a government agency should not produce the record in native format, but instead should produce the metadata in load files[165] that will accompany the static image format of the

---

[159] *See generally*, Cynthia K. Courtney, *Producing Electronically Stored Information in Compliance with the Amended Federal Rules of Civil Procedure*, PLI Order No. 24134 (July/Sept. 2010).

[160] *See supra* notes 65–70 and accompanying text.

[161] It should be noted that a court should try to refrain from compelling the disclosure of certain metadata for reasons mentioned in this Article. *See infra* Part IV. Nevertheless, the disclosure of certain metadata may be compelled if it is not exempt and is of enough genuine public interest to warrant such a disclosure. For example, if certain metadata, such as the author of a document or the time a document was received, is requested by a party, a court may determine that it is of enough genuine public interest to compel such disclosure. In such a circumstance, this model provides a framework for how such metadata should be reproduced by the agency.

[162] *See supra* Part IV.

[163] *See* Krause, *supra* note 28.

[164] *Id.*

[165] A load file is a file that corresponds to a set of static images, such as PDFs or TIFFs "and indicates where individual pages or files belong together as documents, to include attachments, and where each document begins and ends." THE SEDONA CONFERENCE GLOSSARY, *supra* note 6. In addition, "[a] load file may also contain data relevant to the individual documents, such as selected metadata, and extracted text. Load files should be obtained and provided in prearranged or standardized formats to ensure transfer of accurate and usable images and data." *Id.*

document.[166] This would ensure that the requester receives the requisite metadata without the high risk of exposing exempt information. Each page should also be Bates numbered, which is a process of assigning sequential numbers or alphanumeric markings to documents so that each page has a unique identifier.[167] Producing the record in static form with accompanying load files and Bates numbers would ensure that the records produced were still in a readable and usable format by the requesting party,[168] would still disclose the necessary metadata requested, and would avoid the potential consequences that would arise if the documents were produced in native format.[169]

VI. CONCLUSION

---

[166] Addressing the issue of metadata production, *see* Aguilar v. Immigration & Customs Enforcement Div. of U.S. Dep't. of Homeland Sec., 255 F.R.D. 350, 356 (S.D.N.Y. 2008) (citing *Sedona Principles 2d* Principle 12 Cmt. 12b) (internal quotations omitted) (internal citations omitted), where the court found,

> [The] production [of a document] in native form gives the receiving party access to the same information and functionality available to the producing party and requires minimal processing time before production. However, information in native form is difficult to redact or Bates number and the requesting party may not have the software necessary to open the document. By comparison, a production in static image form, such as TIFF or PDF, can be Bates numbered and redacted, but entails the loss of metadata . . . . [I]n an effort to replicate the usefulness of native files while retaining the advantages of static productions, image format productions typically are accompanied by load files, which are ancillary files that may contain textual content and relevant system metadata.

[167] *See supra* note 126 and accompanying text.

[168] *Aguilar*, 255 F.R.D. at 356; Thomas Y. Allman, *Conducting E-Discovery After the Amendments: The Second Wave*, 10 SEDONA CONF. J. 215, 222 n.94 (2009) (quoting *Sedona Principle* 12, Cmt. 12b, Illus. 1) ("[I]maged production with load files satisfies the goals of Principle 12 since in a usable form, *i.e.*, electronically searchable and paired with essential metadata.").

[169] As noted earlier, a government agency will "strive to handle all FOIA requests in a customer-friendly fashion." GSA, *supra* note 54, at 2. However, individuals making requests under FOIA may still appeal to the government agency and then file with the U.S. District Court where the individual lives if the agency denies a request or fails to comply. *Id.* at 9. This ability to appeal provides a safeguard to any abuse or deceit by the government agency.

This Article has shown that the question of whether an agency must produce metadata when requested under FOIA comes with no easy answer. Inherent in the purpose of FOIA is the concept of openness and public access to government information.[170] However, in a world that is contaminated by individuals fraught with ill intentions, it is necessary to find a balance between providing certain information to the public and ensuring the protection of certain information to protect citizens' privacy and national security.

The concept of metadata production in ordinary litigation is becoming increasingly popular as metadata has the potential to be extremely helpful in the discovery process. Despite the potential for contextual inaccuracies, metadata can help determine, among other things, the authenticity of a document. As a result, multiple courts have held that metadata should be produced if requested during discovery under Rule 34 of the Federal Rules of Civil Procedure.[171]

Nevertheless, what has yet to be officially addressed is whether metadata must be produced when requested under the Freedom of Information Act. Only one court has spoken to the issue and in doing so applied the framework of the FRCP to metadata production under FOIA.[172] Despite the opinion being withdrawn on other grounds, the case sheds light on how a court, in the future, may determine whether metadata should be produced when requested under FOIA. Using the FRCP to govern the FOIA process is inherently flawed because FOIA is a separate statute with separate requirements and multiple exemptions within the Act make the FRCP framework impractical. Using the FRCP to govern metadata production under FOIA would allow a court to incorrectly compel the disclosure of certain public records in native format just so the records' metadata remain intact.

As shown, producing records in native format so that all metadata remain intact with a record is impractical and unreasonable under FOIA. To do so would risk the inadvertent disclosure of sensitive information, such as policies regarding military operations or the social security numbers and addresses of individual citizens. In addition, the contextual inaccuracies of metadata can contribute to public misconception and government distrust. Sacrificing the safety of individuals so that every piece of metadata remains intact with a record is irrational.

---

[170] *See supra* note 2 and accompanying text.

[171] *See, e.g.*, Breen, *supra* note 35.

[172] *See supra* note 10 and accompanying text.

As a result, this Article proposes a framework that ensures not only that any requested information is produced in compliance with the purpose of the Freedom of Information Act, but also ensures the protection of sensitive information that, if inadvertently disclosed to the public, could have devastating consequences both to the security of the nation as well as the privacy of the individual citizen. To receive metadata under FOIA, an individual must specifically request what metadata is desired. If the metadata is not exempt under exemption (b)(5) and is readily reproducible, it should be reproduced by the government agency. However, if the records must be produced, they should only be produced in static image format with accompanying Bates numbers and load files. This proposed framework properly balances the two opposing concepts of promoting government transparency and safeguarding sensitive information and needs to be considered before any court or agency decides to produce records in a format that can cause a serious threat to the nation's security or individuals' privacy.