

White Paper: Recommended Format for Electronic Documents

Problem

Administrative Orders 2009-01, 2009-43, and 2009-74 place various electronic court documents in the mainstream of operations from a business perspective and require Bar members to have a valid e-mail address with which to receive case-related communications from the court. Various clerks of court requested clarification of the required format for distribution of court documents after some local Bar members complained that they could not read the .docx format generated/distributed by the case management system and requested that the court send them minute entries in various other formats.

Supreme Court Rule 94 directs clerks to record various court matters and allows for records to be created on a computer or be an image of an original record. The court comment that accompanies the rule specifies that the work must conform to the technical standards approved by the Supreme Court. ACJA § 1-506, *Filing and Management of Electronic Court Documents*, specifies the technical standards for all documents filed or delivered electronically. The code section has been in existence for almost 8 years without any update. It specifies PDF or XML be used for document formatting, but provides for the Supreme Court to adopt further direction regarding standards for XML documents. The section also contains separate format requirements for non-text, graphic documents and multimedia files (TIF, GIF, or JPEG).

Clerks are now requesting direction regarding the “Supreme Court approved” format for electronic documents, since word processors create one format, scanners create another, and filers often protect the format of their documents submitted to the court using yet another. Ushering in e-filing statewide opens the courts to provide and receive electronic documents to and from case parties and counsel, greatly complicating matters if a manageable standard is not clearly communicated quickly. Increases in number of filings will translate into increases in required bandwidth and storage space, so any standard must take into account the related file size overhead that accompanies its use.

Solution

From the technical perspective, size of the image or document is critical. CIO Karl Heckart is very concerned about the total storage footprint made by any electronic document throughout its entire lifecycle. The size of the file has implications for the amount of disk space needed, how much network bandwidth is needed to move the file around, how much backup capacity is needed to preserve the file on various media, and the effort required to migrate to new technology. Court users are concerned that a proprietary format not be adopted that would cause them to have to purchase a specific product in order to file documents with the court or read documents filed by other parties or distributed to them by the court. Another fundamental direction is that document protection is a function of the court’s electronic document management system, not the format of the document itself. In addition, judges need enhanced document usability including text searchability within electronic documents they use (a group of judges is being consulted to formulate specific automation requirements for the bench).

It is also important to take into consideration that 1-506 currently specifies separate requirements for text-based and image-based electronic documents, so one size does not have to fit all.

Taking all these items into consideration has yielded a proposal that an open standard format produced by commonly available word processors be required for text-based electronic documents and the TIFF format be required for images, graphics, and non-searchable documents. Two “rival” XML-based formats approved by the International Standards Organization (ISO) exist: the Open Document Format (OpenDocs) and Open Office XML (OOXML). Microsoft is the main proponent of OOXML, having secured the approval for Word’s particular creation of XML output rather than conforming to the existing OpenDocs ISO specification. OpenDocs is more of a consortium-driven approach subscribed to by Microsoft’s rivals, including OpenOffice.org, Star Office, Lotus Symphony, Google Docs, etc.

It makes sense when adopting an open standard to adopt the “most open” possibility to provide the widest breadth of product selection to users. The court-standard word processor, Microsoft Word 2007 SP2, can read OpenDocs format and save as OpenDocs format, while few other word processors read and save as OOXML. Therefore, the OpenDoc standard has been recommended. The table of file size comparisons illustrates that OpenDocs provides the smallest file in testing among sample, text-based documents of various lengths. While the overhead added by the PDF format seems inconsequential for a single document, when multiplied by the millions of documents that will come through AZTURBOCOURT.GOV each year, even as small size difference proves highly significant.

Recommendations

Revise text of ACJA 1-506(D)(2)(b)

FROM: “XML (Extensible Markup Language), after the supreme court adopts standards for its use.”

TO: “Open Document Format for Office Applications, ISO/IEC 26300:2006 or later.”

Leave remaining text of subsection D as is.

The standard is available from: http://www.iso.org/iso/catalogue_detail.htm?csnumber=43485.

Research Results

File Size Comparison Chart

Format	3 Pages (minutes document)	13 Pages (minutes document)	100 Pages (minutes document)
Word 2007	34 KB	47 KB	180 KB
Word '07 password protected edit*	34 KB	47 KB	180 KB
Open Document Text	16 KB	38 KB	172 KB
Open Document Text password protected edit**	n/a	n/a	n/a
TIFF images (multipage)	122KB	571KB	4.8MB
TIFF Black & White (300 dpi)		743 KB	
TIFF Grayscale (300 dpi)		10.1 MB	
TIFF Color (300 dpi)		12.4 MB	
Pdf image Made with Capture Perfect	111 KB	514 KB	4.39 MB
Searchable pdf (Pre OCR) Made with Capture Perfect	118 KB	516 KB	4.39 MB
Searchable PDF (Post OCR) Made with Capture Perfect	165 KB	717 KB	6.07 MB
Searchable PDF Acrobat distilled	46 KB	103 KB	559 KB

*Protection was applied using Tools drop down at bottom of “save as” dialog box, general options, password to modify (no password to open), protecting against edit but not against being opened, then saving with same filename. This allows editing but not saving. Can also use the Developer Tab, Protect Document function and select “restrict formatting and editing” then limit editing for everyone to make document read only.

** Protection was applied using the same method as above, however the protection did not actually apply when opening the .ODT file. It never prompts for the password, and allows editing if you like. Does not matter if the protection is applied before or after the document is saved in the new format. Also, unable to protect the document either.

Pro's and Con's of Various Format Choices across Complete Document Lifecycle

Format Option	Receipt		Storage		Distribution	
	Pros	Cons	Pros	Cons	Pros	Cons
Word 2007	Reads OpenDocs XML or OOXML, court standard, future focused, could meet 1-506 as is, read by ACAP PCs, can be locked	'07 not large installed base outside courts, version is <u>vital</u> (no XML in pre-'07), locking takes advanced knowledge	CMS generates it and expects to receive/store it, no overhead for locked document. OnBase can store it native, Windows can store it	Easily mixed with court internal documents if not placed in segregated storage	CMS produces, universally read by XML tools, current AJACS process creates	Requires conversion kit download or XML tools, current process doesn't lock
Word 97-2003	Nearly ubiquitous, de facto industry standard, read by ACAP PCs, can be locked	Not XML, not 506 compliant, legal offices late adopters (still love .wpd), locking takes advanced knowledge	CMS can still manage it since earlier version, OnBase can store it native, Windows can store it	Needs read into newer format when XML required	Nearly ubiquitous (except in legal profession!)	Not sent by AJACS CMS, requires extra step to convert and lock
.PDF, basic	Now ISO standard (@V1.7), Locks format (earlier than Acrobat8), even if unlocked characters hard to change & resave	Requires separate program to convert from native format, Two files to save, no longer locks format totally	Meets 506 reqts, can store in OnBase or Windows can store it	Requires conversion from native format, adds overhead to file size	Meets 506 reqts, Nearly ubiquitous, reader is free, locks format (earlier than Acrobat8)	Requires separate step for court to create, format can be unlocked in >Acrobat 8)
.PDF, searchable	Now ISO standard (V1.7), Locks format (earlier than Acrobat8), word processors generally can save to it in current releases	Requires separate program to create, only from a word processed document, can be edited in Acrobat8	Meets 506 reqts, can store in OnBase or Windows can store it	Requires conversion form native format, takes up most storage space of text files	Meets 506 reqts, Nearly ubiquitous, reader is free, locks format (earlier than Acrobat8), more useful than straight pdf	Requires separate step for court to create, format can be unlocked in >Acrobat 8), takes extra resources to create

Format Option	Receipt		Storage		Distribution	
	Pros	Cons	Pros	Cons	Pros	Cons
TIF	Default for scanned documents, protects format in transmission, hardest format to modify and resave, views w/ Windows Fax viewer or other products	Only a picture, not searchable, paper centric, requires local print driver to create	Default for images made from paper stored in OnBase, great for large documents (each page can be separate image)	Doesn't meet current 1-506 text req'ts. Requires metadata to be located, only a picture so not searchable, giant file size, multi-page file clogs network	Only a picture so format locked, std software on PCs (like Windows FAX viewer) reads file type	Requires plug-in to browser. Free readers abound, Adobe owns

Password protection is a function of the EDMS, not the word processor.

Gartner Input

Article: “ISO Standard 'Office' Formats Overpromise Compatibility”

Gartner’s perspective is on enterprise purchase and pricing of next-generation word processors

ODF is a standard. So is OOXML. However, the war over products continues, which is good news for users. Watch which products support which formats, and how converters and other tools supporting compatibility between the two standards evolve. A format that is not tied to a product but tied to a future vision of what products should be and that puts everyone on a level playing field may be the best alternative. However, this scenario is unlikely to happen in the short to midterm.

Longer term, Web-based products likely will represent a more significant threat to Microsoft Office than today's instantiation of OpenOffice.org. If Web-based products are able to combine the rich functions of Microsoft Office, WordPerfect, or even OpenOffice.org, with the management ease and collaboration aspects of Web-based products, then their chosen format, even if that is ODF, could yet be what a standard is supposed to represent.

Article: “Client Computing Survey for 4Q08 Shows Steady Uptake of Office 2007”

Office 2000 support ends in July 2009. Office XP is supported through 2011. In our 2007 North American survey, respondents indicated that 10.7% of users were running Office 2007 — an unusually high percentage for a product that had been shipping for less than a year at the time of the survey. In addition, 2007 survey respondents forecast that a similarly high 41% of users would be running Office 2007 by year-end 2008.

The 2008 survey numbers were more realistic, with 15.1% of users running Office 2007, and a forecast of 47.5% expected to be running it by year-end 2009. Although the U.S. survey does not show a significant increase in the OpenOffice.org (including StarOffice and Symphony) installed base, we continue to receive a significant number of inquiries on this topic, most of which involve augmenting Microsoft Office with OpenOffice.org to cover the number of licenses needed. The installed base of OSO products has been hovering around 2% in these surveys for several years. This is a small number from a relatively small survey, so it's difficult to make any firm predictions, because one or two large respondents can skew the data. This year's survey shows the current installed base of OSO products at 3%, with a forecast showing a doubling to 6% by year-end 2009.

We asked respondents what percentage of documents would be created using which document formats in 2008 and 2009. As expected, the percentage of Office 2007 in use will exceed the percentage of documents created in the new format, meaning that many will continue to use the legacy binary formats. However, the percentage of OOXML documents is still fairly high, and it is likely to represent half the documents created in these organizations in 2010. ODF trails and is likely to continue to do so. If strong OOXML filters appear for OpenOffice.org, then most organizations are likely to opt for OOXML, unless they continue with Microsoft binary formats, which will relegate ODF to use by certain government organizations and open-source purists. Fewer organizations answered this question than the one involving how many users were running each Office product, so this question is less accurate than the others.

Google Docs intrigues organizations, but interest in it continues to exceed implementations, which are user-led.

Article: “Considering Alternatives to Microsoft Office”

A great deal of hype surrounds alternatives to Microsoft Office products. Open-source products — such as OpenOffice.org and its various distributions (including those from Novell, Sun Microsystems and

IBM/Lotus) — have been of interest to organizations since the late 1990s. Other offerings are newer, such as the Web 2.0-style products and services, including Google Docs, Adobe Buzzword, ThinkFree Office and Zoho Office. We have seen minimal deployments of any of these products replacing Microsoft Office in the enterprise.

Article: “Appropriate Formats for Archiving Electronically Signed Documents”

The "format" of the document is a key issue for long-term archiving — not simply archiving electronic content or files but archiving the exact form the document took, what is often called "page fidelity." A document with a proprietary format — Microsoft Word, for example — may be used, but the format in that example is controlled by Microsoft and does not follow national, international or even consortium-based standards. The longer a document created in proprietary format must live, the more likely that there will be software incompatibilities at the time or place when the document must be rendered.

Enterprises that use document formats that require a proprietary reader face increasing risks that the documents will not be rendered accurately over time.

Responses to Specific Concerns Raised

1. **How does clerk filestamping (or automated filestamping in Maricopa) work with the .odt format?** ODT documents accommodate headers/footers and watermarks, the most common methods of applying manual filestamps. A process to apply them can be completed in Word, then the “Save as...” feature employed to change the format to .odt. AOC representatives viewed the Maricopa clerk’s automated filestamping application on Sept. 23. It uses a program to integrate with a database of information about the filing which it applies at the end of the clerk review process (upon acceptance by the clerk). This takes place prior to saving the accepted document in OnBase. All users of the document from that point on see the stamp, including justice partners, public access viewers, and recipients of any official records request fulfilled by the Clerk’s Office. Since the Clerk’s Office converts any document received to .pdf using AdLib prior to clerk review, a change in the format of filed items has no effect. Only a prohibition on conversion of filed documents or a requirement to store filed documents solely as .odt would force a process change for the Clerk. Accepting documents in .odt format would make no difference in the clerk’s current process, since the conversion to pdf occurs prior to application of the filestamp.
2. **How do the process changes impact courts currently requiring other formats?** Older formats will be grandfathered. Individual courts are entitled to convert the formats behind the scenes for storage, should they choose. The requirements primarily cover documents coming into and out of the court, but have been designed to reduce storage requirements to the benefit of all courts.
3. **What training is necessary for public e-filers?** Because great variety exists in the word processing world, training is necessary for all filers regardless of format. Users must be made aware of the court requirements, whether Word, PDF, or OpenDocs. Older word processors will have difficulty outputting the open format but they also don’t build pdfs apart from a second program. Use of older word processors will phase out over time as support for them ends. Users of newer word processors simply need to be shown the specific format to select in the “Save as” dialog box. TurboCourt already has a comprehensive communication and marketing plan. Content such as this falls within that effort. The filing webpage will certainly display help about approved format for free-form filings. Bar members can be informed during e-filing CLE sessions taught by court representatives.
4. **Are non-text format requirements separate from text format ones?** Yes. ACJA 1-506 currently separates the two and acknowledges that exhibits and evidence will not conform to the format of text-based pleadings. The section even contains standards for multimedia files related to court proceedings. These standards are not changing.

- 5. Does every OnBase administrator know how to properly secure documents in OnBase?** The OnBase administrator configures security following implementation. OnBase is an enterprise document management product having robust security capabilities that ensure documents are appropriately protected from view and use. The features figure prominently in administrator training provided by Hyland. Security can be separately configured at each of the following levels: Document type group, Document, and User group. The OnBase administrator assigns individual users as members of function-based groups, decides which documents the groups have access to, and sets up their privileges of use. When a user is a member of multiple groups, the permissions carry the highest level of access granted. At the document type group level, administrators specify which user groups have access to the documents in the group as well as the retention schedule, if either of the OnBase retention modules is employed. At the document level (within a document group), the administrator can override any user group's permission levels for any individual document. Like at the document group type level, the administrator can change whether a user group has rights to retrieve/view a document, create a new document, modify an existing one, delete a document, print a document, mail the document externally or internally, view and/or create renditions and revisions of the document (requires EDM module), view keywords, and modify keywords. As long as the administrator properly configures security, electronic documents will be protected to the level desired by the business process owner.
- 6. Will OnBase allow changes to a text-based electronic document stored on the system?** OnBase has a module, called EDM Services, which allows an administrator control over renditions and revisions of documents. If you have a second rendition of a document, you have the option of setting it as read only. If the EDM Services module is not present, OnBase will import other documents as read-only but will allow a user to change the text. After making any changes in the document, the user is confronted with an import dialog box before closing/saving. OnBase will not save over the original document, but creates a second copy of the document even if the keywords (metadata) are unchanged. If the administrator prevents the user from deleting documents, OnBase keeps a copy of all the related documents and when they entered the system, for audit purposes. OnBase has other internal transaction logs, too.
- 7. What happens to the fillable PDF forms accepted by courts today?** These will be phased out as TurboCourt intelligent forms are made active for each jurisdiction. TurboCourt has selected .pdf for their nationwide standard output format, but they will not be storing Arizona documents.
- 8. How will the transition be handled for cases having mixed format documents, like TIFF and docx.**
Going forward, text documents will be saved in the .odt format while scanned documents and graphics will still be saved in TIFF. Either can be viewed through the OnBase viewer – TIF natively and *.odt as a custom document format using the local path to Word.exe or the Word viewer executable file.
- 9. Maricopa doesn't have Word 2007, so how will they distribute XML?** Older formats will be grandfathered. Maricopa will upgrade their desktop hardware, operating systems, and Office productivity software before XP gets dropped from general support. Even if they are unable to distribute odt format today and have to continue using pdf, they will be able to make the change following their upgrade activities.
- 10. What about the requirement that documents distributed by the court be protected from alteration?** Applying this requirement to copies of a record is common in local practice but not in court rule. The official record of the court stored by the Clerk must remain unalterable. That record resides on the Clerks' EDMS and is protected by security. Copies distributed are not the official record and those desiring an official record must obtain it directly from the Clerk. AOC recommends that justice partners be given access to the official record via the EDMS rather than a distributed document so that no justice partner is required to base their actions on a copy of a record that could have been altered.